Reliability and Validity of the Progress Questionnaire:

An Adaptation of the Outcome Questionnaire

A Thesis

Submitted to the Faculty

of

Drexel University

by

Jason E. Chapman

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

June 2003

DEDICATIONS

For Andressa - Her wonderful support, dedication, and patience over that last five years have made this project feasible. With her, the challenges have been surmountable, the stresses endurable, and the (limited) spare time enjoyable. She has willingly made many sacrifices over these years, and for this I am forever grateful.

ACKNOWLEDGMENTS

There are several individuals who have directly or indirectly facilitated the development and completion of this project. Thanks to Cathy Bolton for providing me with the research consulting opportunity that ultimately lead to the conceptualization of this study and for her assistance with the implementation of the Progress Questionnaire. Thanks to Maureen Hart and Geoff Gray at OQ Systems, Inc., for providing the massive sample of longitudinal data that permitted me to address specific questions about the psychometric properties of the OQ-45.2. Thanks, also, to my dissertation committee (Mike Williams, Ph.D.; Pamela Geller, Ph.D.; Naomi Goldstein, Ph.D.; Ralph Turner, Ph.D., and Elizabeth Turk-Karan, Ph.D.) for their helpful comments and feedback during the preliminary proposal of this study. Finally, thanks to Mac Turner whose teaching initially sparked my interest in statistical analysis and psychometrics, providing me with the solid technical foundation needed for completing this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Reliability and Validity of the Progress Questionnaire:
An Adaptation of the Outcome Questionnaire
Jason E. Chapman
J. Michael Williams, Ph.D.

Patient-focused research is a methodology that involves the regular measurement of patient progress in treatment and the provision of feedback to clinicians to allow modification of interventions to maximize outcomes. A critical component of patient-focused research endeavors is the availability of psychometrically sound assessment questionnaires, and one such measure is the Outcome Questionnaire (OQ-45.2). This investigation was comprised of three studies. Study 1 examined the factor structure and internal consistency of the Progress Questionnaire (PQ), an adaptation of the OQ-45.2, in 278 patients seeking psychotherapy and/or medication management services at a large, urban outpatient mental health clinic. Study 2 examined the factor structure and internal consistency of the OQ-45.2 in a sample of 450 patients receiving outpatient psychotherapy in numerous locations. Study 3 examined the temporal stability of the factor structure of the OQ-45.2.

The PQ and OQ-45.2 were found to possess desirable estimates of internal consistency, similar to those previously reported for the OQ-45.2. The theoretically derived three factor structure of the OQ-45.2 was submitted to Confirmatory Factor Analysis (CFA) and found to be implausible for the PQ and the OQ-45.2. The studies next turned to exploratory procedures so as to investigate the empirical factor structures of the questionnaires. Results of a Principle Components Analysis (PCA) with Promax rotation revealed that the PQ was comprised of ten correlated factors accounting for 60%

of the observed variance. Similarly, the OQ-45.2 was comprised of nine correlated factors accounting for 62% of the observed variance. The sample size in Study 2 permitted validation of this factor structure through CFA. The nine factor model provided significantly improved fit to the three factor model previously tested but was still implausible. Study 3 found that the nine factor structure obtained with baseline data was not stable when imposed on a sample of data from the fourth assessment. The results suggest that the PQ is statistically less effective than the OQ-45.2, and further use of the modified version is not recommended. Possible explanations for the poor fit of the factor structures are offered, and guidelines are provided for future psychometric studies examining the OQ-45.2.

1. INTRODUCTION

Numerous questions pertaining to the delivery and outcomes of psychotherapy have long faced researchers and practitioners within the field of clinical psychology. Examples range from those concerning the benefits of a particular treatment approach for patients, to those concerning which of several treatment alternatives is maximally beneficial to patients with specific diagnoses, to questions of the "active ingredient" of a given approach, to more recent, practical issues such as the authorization of treatment sessions by managed care organizations. Research designs that aim to answer inquiries such as these are both influenced and determined by the unique features, qualities, and considerations associated with the different methodological approaches and the settings in which they occur.

Psychotherapy outcome studies have traditionally been classified as falling between two ends of a continuum, one end occupied by efficacy research methodologies and the other by effectiveness research methodologies (Kazdin, 1999). While the preceding are accompanied by specific limitations, such as the lack of generalizability associated with efficacy research, each approach offers unique advantages that allow for the investigation of therapy outcomes. The use of random assignment, control groups, treatment manuals, and therapist training typify the domain of efficacy research (Dornelas, Correll, Lothstein, Wilber, & Goethe, 1996). These investigations are characterized by an emphasis on the limitation of threats to internal validity, thereby allowing the determination of causal inferences (Kazdin). Because this approach enjoys a high degree of internal validity, randomized clinical trials are often considered the "gold standard" of research methodologies.

Occupying the opposite end of this traditional spectrum of research methodologies is effectiveness research. This approach attempts to capitalize on the limitations of efficacy research, namely lack of external validity. Thus, effectiveness research is characterized by a greater degree of external validity (Goldfried & Wolfe, 1998). Rather than focus on control of numerous threats to internal validity, effectiveness research methodologies focus to a greater extent on the generalizability of findings through investigation of the clinical setting itself, as it occurs outside of the controlled laboratory (Kazdin, 1999). Whereas this approach relies on the collection and post-hoc analysis of data, a newer research methodology, patient-focused research, takes the notion of generalizability a step further through its simultaneous collection and clinical use of the data.

1.1. Patient-Focused Research

According to Lyons, Howard, O'Mahoney, and Lish (1997), while a great deal is currently known about the ability of various therapies to work under controlled circumstances, increasing recent concerns focus more directly on the nature of the relationships among consumers, treatments, and providers. In other words, provided empirically derived knowledge of treatments' abilities to work under ideal conditions, research must then aim to answer questions about treatments as they are routinely delivered in actual clinical practice. In response, a brand of outcome research receiving increased attention of late, and offering benefits unique from the efficacy and effectiveness research methodologies described above, has emerged, one in which the global aim is the improvement of psychotherapy outcomes through increased focus on

and attention to the treatment responses of individual consumers rather than groups of consumers (Lambert, 2001).

While efficacy and effectiveness research studies answer questions about whether treatments are *able* to work or *do work* in applied settings, respectively, the patient focused research approach is concerned with whether a given treatment *is* working for the individual consumer (Howard, Moras, Brill, Martinovichc, & Lutz, 1996). Among the remaining attributes that serve to demarcate patient-focused research methodologies from the traditional methodologies are: the use of statistical analyses, such as hierarchical linear modeling, specifically tailored to its goals and sensitive to individual consumer change; emphasis on cost-effectiveness; aims of quality improvement through modification or adjustment of treatments delivered; increased generalizability of findings; and the provision of feedback to the practitioner (Lambert, 2001).

As may be evident from the characteristics, focus, and aims just discussed with respect to the patient-focused research methodology, a common arena for the conduction of this type of inquiry and investigation is the managed care setting. Several reasons exist for this, including: (a) the rapid growth of managed care and its associated focus on service effectiveness, (b) the growth of consumer-based groups concerned with the quality of services provided, (c) information technology and the ability to efficiently utilize and maintain massive electronic databases, and (d) a new focus based in the application of outcome data for the purpose of refinement of treatments delivered (Lyons, et al., 1997). Likewise, Pallak and Cummings (1994) cite increasing costs associated with service delivery and the availability of applicable assessment instruments, with respect to content and practicality, in managed care settings as additional reasons for the increase in

outcomes assessment observed under managed care. An additional factor in the relationship between managed care and patient focused research is that the position and role of managed care organizations requires decision making with reference to a number of factors associated with therapy. Lambert, Huefner, and Nace (1997) state that such organizations commonly make treatment decisions about the types of treatments to cover, the number of sessions to authorize, and which settings and levels of care to provide.

The conduction of patient-focused research within the managed care setting offers a number of benefits, among the primary of which is access to and utilization of large and comprehensive sets of data about each consumer (Lambert et al., 1997). That is, managed care companies possess sizable, electronic databases containing an extensive amount of routinely gathered information for each consumer. Thus, the authors argue, patient-focused research within the managed care setting bypasses one of the major obstacles of traditional psychotherapy outcome research, sample size, as samples entering into the thousands are frequently observed.

Lambert et al. (2001) describe the development of the patient-focused research methodology as largely the result of economic and practical considerations as well as limitations in traditional outcomes research. Accordingly, two standards for which patient-focused research must strive are cost-effectiveness and increased external validity (Lambert et al.). Though acknowledging the specific, significant benefits of methodological considerations such as multimodal assessment, the settings in which patient-focused research frequently occur cannot endure the time, financial, and other resource burdens required by such strategies. Lyons et al. (1997) argue that if these

limitations are ignored, the consequence is a negative experience for both consumers and providers, resulting in failed research endeavors.

While possessing numerous benefits, given the above constraints, the "gold standard" efficacy research methodology is viewed as impractical and counterproductive in many settings because of its associated resource demands (Lambert et al., 2001). Similarly, effectiveness research investigations, while increasingly generalizable and more consistent with the patient-focused research methodology, are constrained by their focus on group level data and delayed dissemination of research findings. That is, practitioners receive details regarding the effects of treatment subsequent to study completion, thereby preventing alteration and modification of the approach for the purpose of improving outcomes. Additionally, such details are available only at group level, with details specific to individual consumers remaining unknown.

1.2. Outcomes Assessment and the Outcome Questionnaire

In order to achieve the aims discussed above, an effective and efficient means of tracking consumer changes must be developed; however, this raises additional issues: selection of an instrument(s) to use, methods of evaluating the measures, determining measures appropriate for a given setting, and selecting instruments in light of the practical constraints of the setting. According to Lambert et al. (2001) outcome assessment instruments utilized for the purposes of tracking functional outcomes across disorders must be brief, easily administered and scored, suitable across diagnoses, psychometrically sound, sensitive to change across time, and inexpensive.

One such instrument, developed specifically around these aims (Lambert & Finch, 1999), is the Outcome Questionnaire (OQ-45.2; Lambert et al., 1996). A substantial

degree of evidence exists supporting the ability of the OQ-45.2 to meet the criteria listed above (e.g., Lambert et al.). The OQ-45.2 is a measure of the consumer's level of symptoms of psychological distress, interpersonal functioning, and functioning in work, school, or primary role pursuits. The measure consists of 45 items; requires approximately five to seven minutes for completion, so as to be compatible with routine clinical practice; is written at a 6th grade reading level; contains items that are applicable across disorders and likely to be endorsed across consumers; requires minimal oversight by the administrator; requires approximately three to five minutes to score; costs a nominal charge per administration; and possesses sound psychometric characteristics (e.g., Lambert, 1999).

The OQ-45.2 was designed to meet the needs of and provide a valuable resource for two primary groups: managed care organizations and individual practitioners (Wells, Burlingame, Lambert, Hoag, & Hope, 1996). Potential applications of the OQ-45.2 for managed care organizations, described by Wells et al., include: tracking the quality of services delivered, informing decisions regarding treatment allocation, gathering data on the effectiveness of treatments, provider profiling, and with sufficiently large samples, predicting treatment duration, authorizing sessions, and evaluating novel therapeutic interventions. With respect to the needs of the independent practitioner, potential uses include: determining consumer progress, making relevant decisions regarding the course of treatment, providing a baseline measure of pathology, monitoring changes occurring throughout treatment, and assessing the effectiveness of various treatment interventions.

The OQ-45.2 assesses three theoretically derived domains: (a) Symptom Distress (SD), (b) Interpersonal Relations (IR), and (c) Social Role (SR, Lambert & Finch, 1999).

The first subscale, SD, is composed of symptoms representative of the most commonly occurring diagnoses, depression and anxiety. The second subscale, IR, is a measure of interpersonal functioning, including both satisfaction and problem-related items about the consumer's relationships with others. The authors argue that these items were included as a result of the commonly observed relationship, as evidenced in clinical practice, between interpersonal problems and symptoms of distress. The final subscale, SR, is comprised of items relating to the consumer's level of satisfaction with and functioning within his/her primary role area. The items in the SR subscale were selected because, according to the authors, role functioning is commonly impacted by psychological distress and is therefore an important domain in outcome assessment.

The originators of the OQ-45.2 have published several studies detailing findings of the instrument's psychometric properties (e.g., Lambert et al., 1996). Below is a summary of these findings, followed by a discussion of the associated limitations and questions that remain about the operating characteristics of the OQ-45.2.

1.3. Psychometric Properties of the OQ-45.2

Estimates of the test-retest reliability and internal consistency of the OQ-45.2 are based on data collected from two samples: undergraduate students (N=157) and individuals receiving services through employee assistance programs (EAPs; N=290; Lambert, Burlingame, et al., 1996). Internal consistency estimates were calculated separately for the undergraduate and EAP samples. With respect to the undergraduate sample, coefficient alpha for the OQ-45.2 total score was .93, and was .92, .74, and .70 for the SD, IR, and SR subscales, respectively. With respect to the EAP sample,

coefficient alpha for the OQ-45.2 total score was .93, and was .91, .74, and .71 for the SD, IR, SR subscales, respectively.

Test-retest reliability estimates were calculated with data collected from only the undergraduate sample with a three week interval between the two administrations. The overall correlation coefficient was .84, with coefficients for the SD, IR, and SR subscales of .78, .80, and .82, respectively.

Several studies have been conducted for the purposes of estimating the validity properties of the OQ-45.2; the majority of which are subsumed under the category of construct validity investigations. Lambert, Burlingame et al. (1996) used the undergraduate sample described above to assess the convergent validity of the OQ-45.2. The OQ total score and subscale scores were correlated with a number of commonly employed measures of anxiety, depression, social functioning, and interpersonal relationship functioning, so as to assess the convergent validity of the three theoretical domains of the OQ-45.2. The measures used for comparison were the Symptom Checklist-90-Revised (SCL-90-R), Beck Depression Inventory (BDI), State-Trait Anxiety Inventory (STAI), Zung Self-Rating Depression Scale (ZSDS), Zung Self-Rating Anxiety Scale (ZSAS), Taylor Manifest Anxiety Scale (TMAS), Inventory of Interpersonal Problems (IIP), and the Social Adjustment Scale (SAS-SR). Results indicated that the OQ-45.2 correlated highly with the convergent measures. Correlation coefficients calculated between the OQ-45.2 total score and each of the measures ranged from .60 to .88. Among the subscales, the highest correlations were observed between the SD scale and the measures, with correlations ranging from .50 to .89.

Additional support for the convergent validity of the OQ-45.2 was found by Umphress et al. (1997). Results revealed moderate to high validity coefficients between the OQ-45.2 total score and the measures of anxiety, depression, social functioning, and interpersonal relationship functioning, with which it was correlated, ranging from .71 to .84 in the community treatment center sample. Of the three subscales, the SD correlated most highly with the convergent measures, with correlations ranging from .65 to .84, while the remaining subscales produced significant but smaller correlation coefficients. Thus, there is evidence supporting the convergent validity of the OQ-45.2 with commonly employed, theoretically related measures; however, the question remains as to whether the OQ-45.2 empirically is comprised of three subscales or whether it is a global measure of symptom distress.

Lambert, Burlingame, et al. (1996), to assess the construct validity of the OQ-45.2 examined the ability of the OQ-45.2 to demonstrate sensitivity to changes occurring in a consumer sample over time. The authors used a sample of 40 consumers receiving outpatient treatment. Pretest scores were compared to scores after seven sessions of therapy. Results of a repeated measures t test indicated that the two scores were significantly different in the direction of improvement, suggesting that the OQ-45.2 was sensitive to the types of changes resulting from therapeutic intervention. In addition, Vermeersch, Lambert, and Burlingame (2000) used hierarchical linear modeling to provide an additional investigation of the sensitivity of the OQ-45.2 to changes. Results were based on data collected from 1,176 consumers receiving therapy at outpatient mental health centers and a sample of 284 undergraduates who were not receiving treatment. All participants completed the OQ-45.2 on three occasions. Results indicated

that the OQ-45.2 is largely sensitive to change based on the overall score as well as the three subscale scores.

In another article, Umphress et al. (1997) report additional findings in support of the construct validity of the OQ-45.2. This study utilized a sample of community normals (N=210), consumers from a college counseling center (N=53), a community health clinic (N=106), and an inpatient psychiatric unit (N=24). An ANOVA was conducted to determine the ability of the OQ-45.2 to correctly discriminate groups as ranked from least to most severe. Results indicated that means of the patient and community groups differed in the expected directions, from inpatient (highest) to normals (lowest).

The authors reported additional support for the construct validity of the OQ-45.2 from the community clinic sample by investigating whether there were significant differences in the OQ-45.2 score between individuals with Axis I diagnoses and those with V-Code diagnoses. An independent t test supported this hypothesis for the OQ-45.2 total score and two of the three subscales, indicating that those with Axis I diagnoses had significantly higher OQ-45.2 scores than those with V-Code diagnoses.

Finally, Umphress et al. (1997) found that, using the OQ-45.2 total score, they were able to correctly identify 85% of a clinical sample and correctly classify 74% of a normal sample, thus reflecting the sensitivity and specificity, respectively, of the OQ-45.2 when employing its cutoff scores.

Results of a confirmatory factor analysis (CFA) used to investigate the theoretically derived three factor structure described above were less clear (Mueller, Lambert, & Burlingame, 1998). Three models were tested: (a) the three factor solution; (b) a two factor solution consisting of an internal factor, items related to individual

symptom distress, and an external factor, items relating to external events (e.g., relations); and (c) a one factor solution, representing a factor of general symptom distress. While, of the three models, the best fit was achieved by the three factor solution, none of the models demonstrated good fit (Mueller et al., 1998). The authors state that the results "failed to support" the theoretically derived multifactor solution (p. 260).

Thus, as can be seen from that presented above, the OQ-45.2 enjoys a significant amount of empirical support for its psychometric properties. However, this body of research and the associated findings are not without limitations. Mueller et al. (1998) call for additional empirical investigations of the factor structure of the OQ-45.2 and suggest that such studies may necessarily involve modification of the measure. One possible confound in the study presented above, cited by the authors, involves the potential impact of the sample characteristics on their failure to strongly support the three factor structure. The sample was not based totally in clinical populations; rather, it included both community normals (10%) and undergraduate normals (28%). In addition, the clinical portion of the sample contained a large percentage (nearly 50%) from an EAP, a less clinically diverse population. In reality, the community mental health center sample comprised only a small percentage of the total sample (15%), and Mueller et al. acknowledge the potential benefits of limiting future CFAs to an all clinical sample. Finally, the authors recognize the negative impact of their large sample size (N=1,085) on the results of the CFA and the resulting inflation in Type II error rate.

Previous research has found a significant degree of correspondence among the three subscales of the OQ-45.2, providing empirical evidence that it is a global measure of symptom distress (Mueller et al., 1998). However, the results of the study described

above contrast this. Recall that, of the three factor solutions tested, the three factor solution was found to posses the best fit. Thus, contradictory evidence exists regarding the factor structure of the OQ-45.2 from several sources. While some of the existing evidence points to the OQ-45.2 as best characterized as a one factor measure of distress, other empirical evidence exists in support of the three factor solution. As a result, the authors call for further investigation and allow for the possibility of item modification so as to clarify its domains.

Inspection of the correlations among the subscales of the OQ-45.2, the three theoretically derived factors, revealed that a significant degree of interrelationship exists among the three factors (Lambert, 1999). This suggests that, rather than being comprised of three distinct factors, the OQ-45.2 may be better characterized as a global measure of symptom distress (Umphress et al., 1997).

Additionally, each of the studies presented above was conducted by researchers directly involved in the development of the OQ-45.2. That is, none of the empirical findings of the operating characteristics of the OQ-45.2 have been investigated by researchers independent of those directly involved with the measure's development.

The sample characteristics of each of the previously described studies are questionable with respect to generalizability. For instance, approximately 90% of each sample was Caucasian. This characteristic under-represents various ethnic groups routinely observed in clinical practice in many geographical locations. Likewise, the sample for a study by Nebeker, Lambert, and Huefner (1995) on ethnic differences on the OQ-45.2 was comprised of 86% Caucasians. While results were favorable, finding no

ethnic differences, a more diversified sample is desirable, especially when investigating ethnic differences.

Also related to generalizability, a number of the reliability and validity estimates observed above were based on data collected from college undergraduates and/or community normals. The reliability estimates of the OQ-45.2 appear strong according to standard guidelines. Nunnally and Bernstein (1994) suggest that, when scores are used to make decisions about individuals, rather than simply detecting group differences, the guidelines for acceptability of reliability estimates should be increased to .90. Though some estimates were based on clinical samples, those based on undergraduate and/or normal samples may differ in meaningful ways from those observed in actual clinical practice, the setting for which the OQ-45.2 was designed. Thus, many of the conclusions drawn by the authors regarding the desirable psychometric characteristics of the OQ-45.2 are based largely on samples with which the measure was not designed or intended for use and populations which seriously confound conclusions about the measure's reliability and validity properties.

Finally, despite the existence of many studies with large sample sizes conducted on the OQ-45.2, there are no reports of an exploratory factor analysis being conducted, and only one confirmatory factor analysis has been reported. Given the questionable results surrounding the factor structure of the OQ-45.2, as described above, and the numerous large-sample studies conducted, failure to conduct additional analyses of the instrument's factor structure is surprising. According to Nunnally and Bernstein (1994), when empirical evidence fails to support a theoretically derived factor structure, either the relationship was not found, it does not exist, or it exists in a different direction. Thus,

a number of important questions remain regarding the factor structure of the OQ-45.2, and while enjoying empirical support for its operating characteristics, the measure is in need of further investigations and possesses significant shortcomings.

1.4. Modification & Validation of Assessment Instruments

Modification of existing assessment instruments and outcome measures is common practice; this frequently occurs to render a measure more closely suited to the specific purposes and environment for which it is intended and such that it answers the specific questions it is intended to answer (Kazdin, 1993). According to Kazdin (1999), such adaptations, when relevant to a particular setting, are justifiable insofar as the changes are necessary. The benefit of this practice is that when an existing measure is altered, it is not as if one is beginning anew, knowing nothing of the properties of the instrument.

The OQ-45.2 was selected as a means of conducting patient focused research, generally, and monitoring treatment outcomes, specifically. In the current setting the OQ-45.2 has been modified such that it now contains additional items considered to be either missing or underrepresented in the original version. This new instrument has been named the Progress Questionnaire (PQ).The present study is of the type just described; that is, it is a psychometric investigation of an adaptation of the OQ-45.2, the PQ, and will occur in a managed care environment.

The fundamental rationale for proposing this study is twofold: (a) The original OQ-45.2 items are being used in a new environment, thus, as will be shown, necessitating re-evaluation of its psychometric characteristics, and (b) The questionnaire has been altered through the addition of fifteen items, thereby rendering a new instrument. With

respect to the former, the relative degree of validity possessed by a given measure is influenced by a number of factors specific to the environment in which it is employed (Haynes, Nelson, & Blaine, 1999; Nunnally & Bernstein, 1994). Variables such as assessment strategies, target population characteristics, validity judgments, methods of validation, etc. fluctuate from setting to setting and therefore influence estimates of reliability and validity. This argument is echoed by many authors, for instance, Anastasi (1986) argues that because the psychometric characteristics of an instrument fluctuate with respect to both time and setting and are not simply achieved and retained, they must be assessed repeatedly. Likewise, Lyons et al. (1997) state that estimation of a measure's reliability should occur by whomever is conducting research that utilizes the measure. Given the numerous arguments above, the need for psychometric analysis of the PQ operating in a new environment is necessary.

As stated previously, the second reason for the proposed investigation stems from the modification of the original OQ-45.2 through the addition of new questions, rendering a psychometrically new instrument. As a result, evaluation and documentation of the reliability and validity estimates of the new measure is necessary, and there are numerous reasons for the importance of this. According to Kazdin (1999), though some data reflecting the measure's reliability and validity may exist, the new instrument is different, subsequently requiring further demonstration of its psychometric properties. Because there are a number of significant potential uses for outcome data gathered in a managed care environment, ranging from the provision of provider feedback to decision making about session authorizations, psychometric analysis of the new instrument must occur to demonstrate the legitimacy of its use and confidence in decisions resulting from its use.

Outcome measurement is a prerequisite for decision making, and as will be shown, there are additional reasons for investigating an instrument's psychometric characteristics. Indeed, Dornelas et al. (1996) argue that psychometric properties should be among the primary considerations in the selection of outcome instruments. There exists a substantial literature referencing as a required step the examination of the operating characteristics of measures used for the purpose of outcome monitoring (e.g., Berman et al., 1998; & Burlingame et al., 1995). Because the results of such tests serve to inform decisions and consequently affect individuals, the reliability and validity characteristics of the measurements must be considered. Likewise, the Ethics Code of the American Psychological Association addresses test construction and the importance of appropriate procedures for developing and using testing (APA, 1992). Sonnanburg (1996) argues that, despite the purported purpose, observations resulting from any test are unusable and of limited significance unless first critically examined. Burlingame et al. (1995), echoing concern over the use of data from unvalidated measures, state that, "…any decision using the measure is highly suspect" (p. 227). Similarly, lacking test validation efforts, little confidence can be place in the meaning and interpretation of test results (Reckase, 1996). Finally, while Berman et al. (1998) cite the need for the development of sound symptom-related measures, new methodologies for data collection, and the use of large samples Burlingame et al. (1995) suggest that the psychometric characteristics of instruments used for data collection impose an upper limit on the value of any gathered data. Thus, there are numerous arguments for investigations such as this, and following is a discussion of recommendations for performing such investigations.

1.5. Recommendations for Estimating Reliability & Validity

Numerous recommendations exist in the literature as to which types of reliability and validity are to be assessed when conducting psychometric evaluations of assessment instruments. Commonly agreed upon is that this important process is often neglected (e.g., Smith & McCarthy, 1995). Though variable at times, recommendations offered were largely consistent in scope. Following is a brief discussion of the types of reliability and validity most frequently evaluated and reported in psychometric investigations such as that presently proposed. These guidelines provide the foundation for the statistical methods described later.

According to Smith and McCarthy (1995), the proper refinement of outcome measures is a two phase process. The first phase of this process involves the identification of the instrument's factor structure and estimation of its internal consistency. Subsequent to this, the second phase entails demonstration of the degree of relationship between the instrument and other important variables. The present study focuses on the former phase, though each phase is briefly reviewed below.

With respect to evaluation of an instrument's reliability, internal consistency is defined as the degree to which the items comprising a measure covary in one individual taking the test one time (Strube, 2000). While internal consistency refers to the degree to which items comprising a test are interrelated, it does not indicate that the test is necessarily unidimensional (Schmitt, 1996). Determination of the reliability of measures is important because it allows for generalization of the results obtained by the measure, and without reliability, validity cannot be established (Nunnally & Bernstein, 1994). Likewise, without establishing the internal consistency of a measure, the relative degree

of error variance of the measure is unknown, thus making uncertain the degree to which an individual's score represents a true index of the measure (Smith & McCarthy, 1995).

Nunnally and Bernstein (1994) state emphatically the importance of internal consistency estimate calculation for new measures or new uses of existing measures, "It [Chronbach's coefficient alpha] is so pregnant with meaning that it should routinely be applied to all new tests," and "…there is no excuse for not computing it [Chronbach's coefficient alpha] for *any* new measure" (p. 234; emphasis original).

Temporal stability, also known as test-retest reliability, represents the correlation between the scores of the same test administered to the same individual with a delay between administrations (Kazdin, 1998). It is important to assess the degree to which scores on the measure are consistent with a delay between administrations, as it provides information regarding the extent of natural fluctuation in scores that is expected over time (Sonnanburg, 1996). This provides an estimate of the stability of the domains assessed by a measure (Groth-Marnat, 1990), though it should be noted that with state-like measures (i.e., symptom focused measures), temporal stability may be lacking in information as a result of naturally occurring fluctuation in scores (Nunnally & Bernstein, 1994).

With respect to the components of validity, while content validity is commonly recognized as one of the principle types of validity (e.g., Nunally & Bernstein, 1994; Smith & McCarthy, 1995), it will not be specifically addressed in the present study for three primary reasons. First, because this investigator was not directly involved in the item selection process, it is inappropriate to comment on the associated decision making process and theoretical rationale without an empirically based approach of doing so. Second, this process, as it relates to the original items of the OQ-45.2 has been described

elsewhere (e.g., Wells et al., 1996). Finally, discussion is deferred based on the recommendations of Nunnally and Bernstein (1994) who suggest that the description of a measure's content validity is limited in its empirical backing and is largely dependent upon argument and theoretical agreement. Thus, the following is limited to discussion of construct validity.

The second major type of validity is construct validity. Construct validity is defined as whether an instrument, in practice, measures that which it purports to measure (Bryant, 2000). Encompassed by the category of construct validation are two additional types of validity: convergent and discriminant (Nunnally & Bernstein, 1994). Respectively, convergent and discriminant validity represent the correlation between the measure of interest and another theoretically similar measure with which it is expected to correlate highly and the correlation between the measure of interest and another theoretically dissimilar measure with which it is not expected to correlate (Kazdin, 1998). Assessment of convergent and discriminant validity provides significant detail about the characteristics of the measure of interest, that it relates in expected ways to similar measures and to measures to which it should not relate. Nunnally and Bernstein (1994) echo this sentiment, recommending empirical investigation of both convergent and discriminant validity, and Chronbach (1990) argues that estimation of an instrument's convergent and discriminant validity properties is one of the key steps in the process of establishing a measure's construct validity.

1.6. Contribution

The current study involves a psychometric analysis of the OQ-45.2 adaptation, estimating various components of its reliability and validity properties. Before turning to

a discussion of the specific methodological considerations in the current study, the potential contribution of that which follows will be addressed. As discussed thus far, investigations such as this are crucial to the proper development and ethical application of new assessment instruments. However, such studies provide more extensive benefits than the simple provision of empirically based evidence for the measure's psychometric characteristics.

With the rise of managed care's requirement of provider accountability, outcome measures such as the OQ-45.2 and PQ will find increasingly widespread applications and play an important role in decision making and quality improvement (Johnson & Shaha, 1996). This is evidenced, in the current setting, by the fact that additional managed care providers will begin implementation of the PQ in the future for purposes similar to those described herein (M. Hart, personal communication, July 5, 2001). Along similar lines, Clark and Watson (1995) argue that because of the real-world applications of many outcome measures, clear demonstration of the operating characteristics of such instruments is increasingly being required. Thus, such an investigation holds the potential to further establish a potentially widely applicable assessment instrument that is largely compatible with current trends in managed care and patient focused research.

The current study is also in response to Mueller et al. (1998) who called for additional psychometric investigations into the factor structure of the OQ-45.2. Firstly, following the recommendation of Reckase (1996), this study will provide a complete review of an assessment instrument by individuals other than those who created the measure. While a considerable amount of psychometric data exists in support of the OQ-45.2, the majority, if not all, of it descends directly from researchers involved with its

development. Secondly, this study will attempt to correct for the previously cited shortcomings of existing research on the OQ-45.2, namely, the inclusion of a large percentage of normal participants for whom the test is not intended. In addition, it will utilize an appropriate sample size for factor analytical procedures.

1.7. Goals & Hypotheses

In this investigation, there were four primary questions of interest which were addressed in three distinct studies. In Study 1 (Chapter 2), the psychometric properties of the PQ were established using an all-clinical outpatient sample. More specifically, the validity of the theoretically defined factor structure of the OQ-45.2 was tested using Confirmatory Factor Analysis (CFA) in the sample of PQ data. In addition, estimates of internal consistency were calculated and expected to be consistent with those existing in the literature for the OQ-45.2.

In Study 2 (Chapter 3), the psychometric properties of the OQ-45.2 were evaluated using an all-clinical sample of outpatient data. Similarly to Study 1, this was accomplished through CFA. The internal consistency of the OQ-45.2 items was estimated, and expected to be consistent with previous findings. In addition, the measure's test-retest reliability was calculated across varying time intervals between administrations. The temporal stability was also expected to be consistent with previous findings.

The purpose of Study 3 (Chapter 4) was to further the aims of Study 2 by examining the temporal construct stability of the factor structure of the OQ-45.2. That is, as an indicator of construct validity, the factor structure of the OQ-45.2 was examined at baseline and fourth administration to determine whether the two are consistent. It was

expected that the underlying factor structure of the measure would be consistent from the first to the fourth administration. While this question has not been specifically addressed in prior literature, it is considered to be an important aspect of construct validity. Because data collected with the OQ-45.2 may ultimately be utilized to make decisions about consumers' treatment, it is important to address this issue to ensure the construct validity of the OQ-45.2. This was accomplished through CFA.

Finally, the fourth aim of the study was to qualitatively evaluate the implementation usability of the PQ in an outpatient clinical setting. As previously discussed, the overall aim of patient-focused research is the ongoing measurement of patient progress with minimally increased burden on office staff and providers. The degree to which this aim was accomplished is discussed in Chapter 5.

1.8. Summary

The patient-focused research methodology was discussed in light of efficacy and effectiveness research approaches. This methodology was applied in a managed care setting, the method of outcome evaluation for which was an adaptation of the Outcome Questionnaire, the Progress Questionnaire. The present study is a psychometric investigation of the PQ. While the OQ-45.2 possesses desirable reliability and validity characteristics, studies conducted thus far possess significant limitations. In addition, the importance and requirement of re-evaluation of an instrument's operating characteristics subsequent to item modification has been discussed. Thus, and consistent with the recommendations of several authors, the primary aim of this study is demonstration of the reliability and validity characteristics of the PQ in an all clinical sample. More

specifically, this entails estimation of internal consistency, test-retest reliability, and estimation of construct validity through confirmatory factor analysis.

2. RELIABILITY & VALIDITY OF THE PROGRESS QUESTIONNAIRE

2.1. Method

2.1.1. Participants

The total sample (N = 278) consisted of new patients presenting to the outpatient clinic at Friends Hospital in Philadelphia for psychotherapy and/or medication management services between September, 2001, and December, 2001. Participants completed a demographics questionnaire along with the PQ. Demographic data were not available for a substantial portion of the sample as this questionnaire was not included in the original group of intake packets. In addition, some demographics questionnaires contained missing data. Table 1 presents the available sample descriptive statistics for participant age, sex, and race.

Table 1: PQ Sample Demographics

Age (n = 234)	
M	39.61
SD	13.07
Minimum	18
Maximum	81
Sex (N = 278)	**n (%)**
Male	107 (38%)
Female	149 (54%)
Unavailable	22 (8%)

Table 1: PQ Sample Demographics (*continued*)

Race (N = 278)		
Caucasian	87	(31%)
African-American	44	(16%)
Hispanic	13	(5%)
Asian/Pacific Islander	4	(1%)
Other	2	(.7%)
White-Native American	2	(.7%)
Unavailable	127	(46%)

Descriptive statistics pertaining to patient diagnosis are presented in Table 2. Diagnoses were available for 194 patients (70%); however, it should be noted that this data is of limited utility as a substantial proportion (30%) of these are deferred diagnoses. This is the result of the guidelines of the Managed Behavioral Healthcare Organization (MBHO), which permit clinicians to assign a diagnosis of 799.9, Diagnosis Deferred, at the initial visit.

Table 2: PQ Sample Primary Diagnosis

Diagnosis	n	(%)
Depressive Disorder	83	(30%)
Diagnosis Missing	83	(30%)

Table 2: PQ Sample Primary Diagnosis (*continued*)

Diagnosis	n (%)
Diagnosis Deferred	83 (30%)
Anxiety Disorder	20 (7%)
Schizophrenia	5 (2%)
Substance Use Disorder	4 (1%)

2.1.2. Measure

As previously stated, the measure evaluated in this study was the Progress Questionnaire (PQ; See Appendix A). This measure is based predominantly on the Outcome Questionnaire (Lambert et al., 1996), and there are currently no psychometric data available on the PQ. The PQ contains fifteen new items for a total of 60 items and requires approximately 7-9 minutes for completion. The PQ is a paper-and-pencil questionnaire completed by darkening a response choice circle with a pencil. The responses were entered into an electronic database for data analysis by hand. Items 1-45 of the PQ are identical to the items of the OQ-45.2 in order, wording, and format. Items 46-60 were added by the MBHO. These items are of theoretically consistent content to the items of the OQ-45.2, possess the original scale, and were deemed by the MBHO to be significant additions to the existing OQ-45.2 items because they increase the breadth of diagnostic, symptom, and functioning-related items considered theoretically consistent with the existing factor structure of the OQ-45.2.

Instructions for the PQ are identical to those of the original OQ-45.2 and ask the participant to respond to the items, looking back over the last week. In the instructions,

the stated purpose of the questionnaire is to "help us understand how you have been feeling."

All items of the PQ are answered on a five point scale identical to that of the OQ-45.2: 0=Never, 1=Rarely, 2=Sometimes, 3=Frequently, and 4=Almost Always. Examples of original OQ-45.2 items include: "I feel nervous," "I am satisfied with my relationships with others," "I feel stressed at work/school." Examples of the new PQ items include: "I have thoughts of hurting other people," "I see and hear things that other people do not see or hear and that might really not be there," and "I have panic or anxiety attacks that come suddenly out of the blue."

The number of missing items allowable on the PQ before declaring an invalid protocol was set at 7 or more missing items. This was based on scoring guidelines for the original OQ (Maureen Hart, personal communication, September 6, 2002) which permits five or less missing responses. Two additional missing items were allowed for the PQ due to the addition of 15 new questions.

2.1.3. Design & Procedures

This investigation was a unique evaluation of the PQ that would otherwise have not been conducted. That is, the purpose, rationale, hypotheses, and analyses of this study were proposed by this author to the quality improvement department and research committee at the MBHO. In the absence of this study, the psychometric characteristics of the PQ would not have been estimated, examined, or reported.

The global aim of the procedures outlined below was to allow for thorough investigation of the psychometric properties of the PQ as used in actual outpatient clinical

practice. Data collected were used to evaluate the factor structure and other estimates of reliability and validity of the PQ.

Individuals who sought services at the site described above completed the PQ as part of the standard intake paperwork immediately upon arrival to their first session. Questionnaires were provided to all patients by front desk staff as part of the standard new patient procedure. That is, participants arrived to their initial appointments and completed the questionnaire as one component of a standard packet of intake documents. The forms were returned to the receptionist immediately upon completion and prior to the appointment. This represents the baseline, and only, administration of the questionnaire to this sample.

2.2. Results: Psychometric Properties of the Progress Questionnaire

The analyses presented below are based on the sample described above. That is, baseline data for 278 patients presenting for outpatient treatment during the data collection period were examined. Because the overall aim of this investigation was psychometric evaluation of the PQ in the context of a patient-focused research project, descriptive data pertaining to the logistical aspects of the questionnaire are important as they provide some indication of the usability of the measure. Such results assist in the identification of areas for further questionnaire, staff training, and patient-focused research development. Presented below are the PQ descriptive statistics.

2.2.1. PQ Descriptive Statistics

Questionnaire completion. The PQ was provided to a total of 371 new patients. Of these, 75% (n = 278) were useable for analysis. The remaining 25% (n = 93) were not useable for at least one of three reasons. In 54% (n = 50) of unusable cases, office staff

inadvertently provided patients under the age of 18 with the PQ as part of the standard intake packet despite their minor status. In 40% (n = 37) of unusable cases, the questionnaire was deemed invalid on the basis of missing data. Finally, in 6% (n = 6) of unusable cases, complete data were not available due to one or more pages missing from the PQ. Thus, the final sample submitted for analysis was comprised of 278 usable cases.

Data entry integrity check. All data were entered into a Microsoft Access 2002 database and then converted into an SPSS format data file for analysis with SPSS 10.0.5 and AMOS 4.0. Based on the recommendation of Kazdin (1998), the accuracy of entered data was first assessed by selecting a random sample of 20% (n = 56) of entered cases and manually comparing the values entered into the database with those recorded by the patient on the paper questionnaire. Of 3,360 data points (56 PQ's with 60 items each), two items (.06%) were entered incorrectly. Thus, of 56 cases, 54 (96%) had all data entered correctly. Based on this finding, no additional cases were selected for data entry accuracy checking.

With respect to the database, it should also be noted that nine PQ items are reverse scored. This transformation was directly implemented during the data entry stage by associating the response item with the reverse score. Therefore, in subsequent analyses and interpretation, all items are coded in the same direction, such that a high score on any item reflects greater symptomatology or impairment in functioning. For example, after being reverse scored, a high score on PQ01, "I get along well with others," reflects increasing relationship difficulty.

Missing data. Inspection of the PQ data revealed a substantial amount of missing data. A total of 376 data points (2.25% of all possible responses) were missing. The

average number of missing items on each PQ administration was 1.35 (SD = 2.22) of 60. A total of 126 (45%) cases had at least one missing data point. Of cases with missing data, the majority, 98 (78%), were missing four or fewer items.

Inspection of the most frequently missing PQ items revealed similar question content. Table 3 contains these items and the percentage of cases in which each is missing. It appears that the most frequently missing items are those that do not apply to some patients, in contrast to items such as "I like myself" or "I feel blue" that is readily applicable to all patients. It should also be noted that these items were, on multiple occasions, accompanied by written notes on the questionnaire such as "N/A" or for work related items, "I don't work."

Table 3: Frequently Missing PQ Items

PQ #	Question	% Missing (n)
07	I feel unhappy in my marriage/significant relationship	12.2% (34)
48	I have crises where I call my therapist	10.1% (28)
17	I have an unfulfilling sex life	9.4% (26)
57	I am taking my medication exactly as prescribed	9.4% (26)
12	I find my work/school satisfying	8.6% (24)
39	I have too many disagreements at work/school	7.2% (20)
38	I feel that I am not doing well at work/school	6.8% (19)
04	I feel stressed at work/school	6.5% (18)
28	I am not working/studying as well as I used to	6.1% (17)
44	I feel angry enough at work/school to do something I may regret	5.4% (15)

Table 3: Frequently Missing PQ Items (*continued*)

PQ #	Question	% Missing (n)
50	I have days where I am less productive at work	5.0% (14)
14	I work/study too much	4.7% (13)

Because the original OQ-45.2 was written and coded such that a response of zero indicates "Never" or "Does Not Apply," the above items were recoded from missing to zero. Following this, none of the 278 questionnaires had greater than five missing items. This leaves a total of 74 missing data points (.4%).

Gorsuch (1983) recommends the use of a regression-based method for the estimation of missing data points. The SPSS Missing Value Analysis Regression subcommand was used to estimate missing items through multiple regression equations (SPSS, 1997). Though AMOS is equipped to handle missing data through its own procedure (Byrne, 2001), the decision was made to utilize regression based imputation so as to provide a complete and consistent dataset to be used in other analyses. Thus, following this procedure, a complete set of data was available for analysis.

2.2.2. PQ Internal Consistency

Cronbach's Coefficient Alpha was computed using SPSS Reliability Analysis to assess the internal consistency of the PQ. Results for the PQ total score as well as the three theoretically derived subscales (SD, IR, and SR) are presented in Table 4.

Table 4: PQ Total and Subscale Internal Consistency Values

PQ Scale	Internal Consistency (N = 278)
PQ Total	.95
Symptoms Distress (SD)	.93
Interpersonal Relationships (IR)	.80
Social Role (SR)	.77

2.2.3. PQ Three Factor CFA

The purpose of this analysis was to provide an empirical test of the plausibility of the theoretically derived three-factor structure of the modified version of the OQ-45.2, the PQ. The three-factor model tested is described below.

Variables removed from analyses. Three of the fifteen new variables were removed from subsequent analyses on theoretical grounds. Each item was deemed theoretically inappropriate for the PQ due to its presupposition of current psychiatric and/or psychotherapy services. These items were: PQ48 (“I have crises where I call my therapist”), PQ56 (“I take medications for mental health problems”), and PQ57 (“I take my medication exactly as prescribed”). In addition, PQ56 is inconsistent with the scaled response choices due to its dichotomous nature.

Conceptual model. This is a confirmatory factor analysis model achieved through structural equation modeling. As previously noted, the OQ-45.2 was constructed with three theoretically derived subscales. The SD subscale is comprised of the following items: PQ02, PQ03, PQ05, PQ06, PQ08, PQ09, PQ10, PQ11, PQ13, PQ15, PQ22, PQ23, PQ24, PQ25, PQ27, PQ29, PQ31, PQ33, PQ34, PQ35, PQ36, PQ40, PQ41, PQ42, PQ45,

PQ46, PQ47, PQ49, PQ51, PQ52, PQ53, PQ54, PQ55, PQ58, PQ59, and PQ60. The IR subscale is comprised of the following items: PQ01, PQ07, PQ16, PQ17, PQ18, PQ19, PQ20, PQ26, PQ30, PQ37, and PQ43. The SR subscale is comprised of the following items: PQ04, PQ12, PQ14, PQ21, PQ28, PQ32, PQ38, PQ39, PQ44, and PQ50.

According to the model tested below, the PQ is a measure of three latent constructs, SD, IR, and SR, each represented by ellipses in the path diagram. Each construct is defined by a number of observed indicators, as represented by rectangles in the path diagram (See Figure 1). The relationship between the construct and the observed variable is represented by a straight arrow. Additionally, the model specifies that all pairs of the three latent constructs are correlated. Finally, associated with each observed indicator is an error term, represented by an ellipse, which represents the unique variance associated with each question. Thus, this is a conceptual representation of the theoretically derived factor structure of the PQ, and subsequent analysis will provide a measure of the extent to which the sample data support this model.

Statistical model. Considering the above model at a statistical level, there are a total of 57 observed indicator variables (PQ01-PQ60, with PQ48, PQ56, and PQ57 removed), 60 unobserved variables (the three latent constructs and the 57 unobserved error terms associated with the PQ items), for a total of 117 variables in the model. This yields 57 free parameters and 60 fixed parameters. Thus, the χ^2 test for the overall goodness of fit of this model is based on 1653 sample moments, 117 parameters to be estimated, and 1536 degrees of freedom.

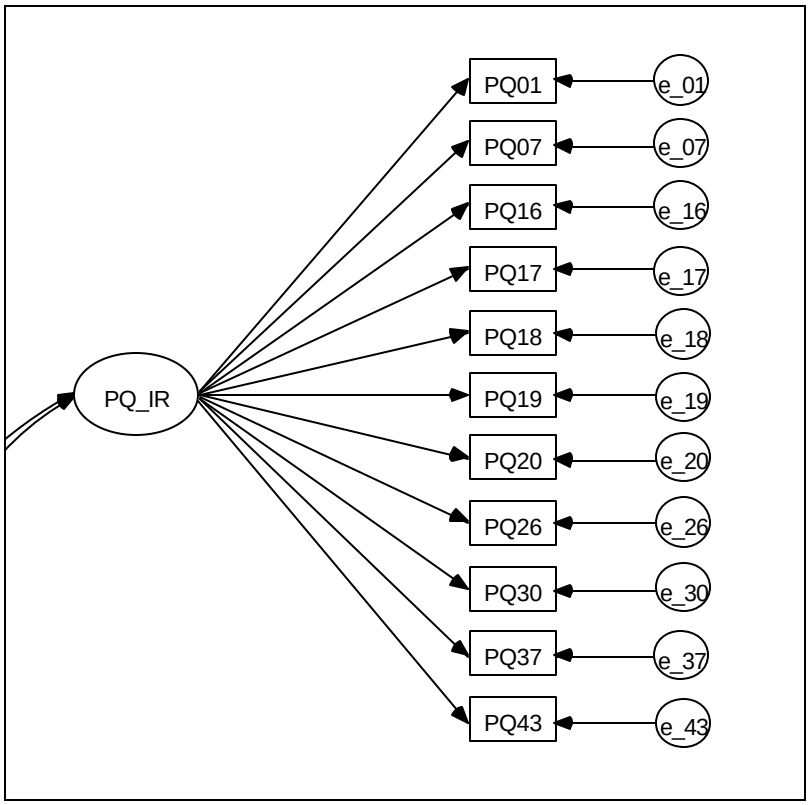

Figure 1: Path Diagram of the PQ Interpersonal Relationships Factor

Matrix to be analyzed. Based on current practices in SEM, the following analysis was based on the covariance matrices of the observed data. Because covariances do not represent meaningful units for visual inspection, the bivariate correlation matrix and standard deviations for all PQ items are frequently reported. However, due to the large number of variables in the present analysis, this information has been omitted and is available upon request.

An additional issue relating to the practice of SEM with psychological data is the controversy surrounding the typical treatment of categorical variables as continuous for analyses based in structural equation modeling. While debate exists, based on the

recommendations of Byrne (2001), as well as standard practice, the data in the analyses below were treated as continuous.

Distribution of data. An important assumption in SEM analyses, such as CFA, is that the data to be analyzed follow a multivariate normal distribution (Kline, 1998). There are two primary methods of making this determination. First, examination of univariate normality estimates, namely skewness and kurtosis, provides details of the distribution characteristics of each variable. The existence of individual variables departing significantly from a univariate normal distribution prohibits the existence of multivariate normality (West et al., 1995). In other words, the condition of multivariate normality presupposes univariate normality. Second, the condition of univariate normality does not guarantee multivariate normality. Thus, given univariate normality, conditions of multivariate skewness or kurtosis may still exist.

Univariate estimates of skewness and kurtosis were obtained through SPSS Descriptives and, the variables with significant departures from normality are summarized in Table 5 below. Kline (1998) argues that skewness values in excess of ±3 and kurtosis values in excess of ±10 represent significantly non-normal univariate distributions. Likewise, he recommends that kurtosis values in excess ±20 represent severe departures from univariate normality. Accordingly, of the 60 PQ items, two variables had significant levels of univariate skewness and kurtosis. The items are: PQ11 ("After heavy drinking, I need a drink the next morning to get going") and PQ32 ("I have trouble at work/school because of drinking or drug use").

These variables were transformed so as to decrease their departures from univariate normality. As recommended by Kline (1998) and Hair et al. (1995), the

selected transformation, provided the degree of positive skew, was 1/(X+1). See Table 5 for the original and transformed skewness and kurtosis values.

Table 5: Non-Normally Distributed PQ Items

PQ Item	Skewness	Kurtosis
PQ11	3.10	10.35
PQ32	3.27	10.78
PQ11 Transformed	-1.19	4.99
PQ32 Transformed	-1.47	3.54

Multivariate kurtosis was assessed by Mardia's test of multivariate normality which was provided by AMOS (Byrne, 2001; Mardia, 1970). Various sources suggest that large values, e.g., those greater than 30, should be considered significant departures from multivariate normality (see Newsom, 2001; Byrne, 1995; and Byrne, 2001). Mardia's measure of multivariate kurtosis was 296.581, the critical ratio of which was 30.148. Based on the above guidelines, this indicates that the PQ data are significantly multivariate non-normal.

Model estimation and fit criteria. The model parameters in the following analyses were estimated through the Maximum Likelihood (ML) method. This decision stems from the recommendations of multiple authors for most analyses based on SEM (e.g., Chou & Bentler, 1995; Hoyle & Panter, 1995). There are numerous reasons for this decision, among them, ML estimation is frequently employed in SEM analyses, has

demonstrated robustness with respect to violation of the conditions of normality, and possesses reasonable sample size requirements (Chou & Bentler).

Below is a brief presentation of the fit indexes utilized in these analyses for the purpose of determining the adequacy of model fit. This includes the definition of each and explication of the critical values against which each was evaluated. There are two basic categories of fit indexes, absolute fit indexes and incremental fit indexes (Hoyle & Panter, 1995). The former is a measure of the degree to which a pre-specified structural model reproduces the sample data. Accordingly, they can be considered measures of the "badness of fit" of a particular model (Hoyle & Panter, p. 165). Conversely, the latter is an indication of the model "goodness of fit," or the extent to which the specified model provides improved fit when compared to a model with no relationships among the variables (Hoyle & Panter, p. 165). The authors recommend reporting at least one absolute fit index along with two or more incremental fit indexes.

The Chi-Square was used as an absolute index of fit. The chi-square statistic tests the lack of fit of the theoretical model tested. Therefore, rejection of the null hypothesis ($p > .05$) is interpreted as a model that does not fit the observed data well, that is, a non-plausible model. However, it should be noted that a known limitation of this test is that the χ^2 is quite sensitive to sample size and assumes perfect model fit, frequently resulting in rejected models (Byrne, 2001).

The most consistently performing indicator of absolute fit is the goodness of fit index (GFI). The GFI represents the relative amount of variances and covariances expected by the implied model that is accounted for by the observed data. Thus, the value of the GFI is similar in nature and interpretation to the value of an R^2 value in regression

analyses (Hoyle & Panter, 1995). The maximum value of the GFI is 1, though the index can be less than 0. A value of one indicates that the observed variances and covariances are identical to those anticipated by the theoretical model. In this investigation, a GFI cutoff score of .90 was used to identify a plausible model, with values below this cutoff suggesting an implausible model.

The Comparative Fit Index (CFI) was also used to evaluate model fit. The CFI is based on an adaptation of the normed fit index (NFI) that considers the influence of sample size when comparing the hypothesized model to the independence model (Byrne, 2001). Based on the recommendation of Byrne, a cutoff value of .95 was used to consider the model well-fitting.

The Tucker-Lewis Index (TLI) was also used to evaluate model fit. The TLI is similar in interpretation to the previous measures of goodness of fit presented here (Byrne, 2001). A cutoff score of .95 was used to consider the model well-fitting.

Parameter estimates. According to Byrne (2001), the CFA results should initially be examined so as to determine the feasibility of obtained parameter estimates. Likewise, adequate parameter estimates can be identified by statistically significant critical ratios, i.e. C.R. in excess of ±1.96, for the regression weights, covariances, and variances (Byrne). Removal of any variables falling in the non-significant range is recommended. Examination of the parameter estimates revealed that all critical ratios were statistically significant. In addition, as is desirable, visual inspection of the standardized regression coefficients revealed no Heywood cases (i.e., coefficients greater than 1 or negative variances). All variables were retained for further analysis. Parameter estimates are included in Appendix B.

Because of the excessively kurtotic nature of the data, the bootstrapping procedure was employed as recommended by Byrne (2001). The bootstrap procedure involves the selection random samples with replacement from the sample data, yielding a sample-specific empirically-based normal sampling distribution (West et al., 1995). Parameter estimates are based on this distribution in the presence of violated SEM assumptions, such that the accuracy of the parameter estimates is markedly improved. The parameter estimates obtained through the bootstrapping procedure are omitted due to space considerations and are available upon request.

Evaluation of model fit. The overall test of model fit indicated that the theoretically derived three factor model was not plausible based on the observed data, $\chi^2(1536, N = 278) = 4028.393$, $p = .0001$, GFI = .616, TLI = .643, and CFI = .656. Given the excessively kurtotic nature of the data, the χ^2 was also calculated according to the Bollen-Stine bootstrap method (Byrne, 2001), $p = .002$. Thus, even when accounting for the non-normality present in the data, the model tested remains implausible.

Model modification. Though AMOS provides two sources of information that can be utilized to assess model misspecification and make decisions regarding the respecification of model parameters (i.e., standardized residuals and modification indexes), the model tested above was not modified in this analysis. This decision was made on the basis of multiple authors, and the primary reasons are briefly addressed below.

First, the exceedingly poor fit of the three factor model just tested suggests that the entire model should be respecified. This falls beyond the scope of modification indexes. Second, modified models run the risk of capitalizing on the sample-specific

characteristics of the data matrix (Hoyle & Panter, 1995). Thus, they must be tested on a distinct sample to account for these concerns; however, the sample size of the present study does not permit the execution of this procedure. Similarly, substantive changes to the specified model should be based in theory rather than empirical indicators of fit, as a purely empirically driven modification is unlikely to yield an accurate model (Kline, 1998). In this case, the theoretical basis for the three factor structure appears to be inadequate and no other theory is offered. Another reason provided for not modifying the model is that, with the exception of very large sample sizes, the modifications are unlikely to replicated (MacCallum, Roznowski, & Necowitz, 1992).

Given the poor fit of the sample data to the theoretical model just described, the focus shifts to an examination of a more appropriate factor specification for the PQ. However, it must be noted explicitly that, at this stage, the analyses change from a confirmatory approach to an exploratory approach (Byrne, 2001). That is, beyond this stage of analysis all results are considered exploratory as the analyses are not driven by theory. The PQ sample data were submitted to an exploratory factor analysis. Results are presented below.

2.2.4. Exploratory Derivation of the PQ Factor Structure

A principle components analysis (PCA) with Promax rotation was performed using SPSS Factor Analysis. The oblique rotation method was selected on the basis of recommendations by Hair et al. (1995) and Gorsuch (1983), who argue that oblique methods are entirely appropriate when the purpose of the analysis is identification of a factor structure in which the latent factors would be expected to be correlated. This is the case in the present investigation.

Prior to conducting a PCA, the bivariate correlation matrix was inspected for correlations greater than or equal to .30, as guidelines suggest that a moderate portion of the correlations should fall in this range (Hair et al., 1995). Inspection of the correlations revealed that approximately 31% (i.e., 547 of 1770) were at least .30 in magnitude. When considering the entire correlation matrix, the average correlation was .31.

In addition, residuals were examined in order to determine the amount of correlation remaining between variables when the effects of the other variables removed (Hair et al., 1995). Ideally, this is a small value, suggesting a high degree of intercorrelation among the variables. Examination of the residuals revealed that 23.5% are greater in absolute value than .10, 3.4% of which were greater than .20, and .6% of which were greater than .30. Thus, with the effects of the other variables are removed, only a small portion of variables retain a significant bivariate relationship. This is a desirable condition for factor analysis.

The Measure of Sampling Adequacy (MSA) values, an additional indication of the extent of intercorrelation among variables and the suitability of the data for factor analysis, were also examined (Hair et al., 1995). The authors recommend that variables with MSA values falling in the unacceptable range (i.e., below .50) should be excluded from further analysis. Examination of MSA values revealed that one variable, PQ57 ("I am taking my medication exactly as prescribed"), fell in the "unacceptable" range with a value of .489. In addition, PQ56 (I am taking medication for mental health problems), with an MSA value of .544, fell in the "miserable" range. One variable, PQ14 (I work/study too much), fell in the "mediocre" range, .617. Six variables fell in the "middling" range, and the remaining variables all fell in the "meritorious" or

"marvelous" range, .80 and above or .90 and above range, respectively. Thus, PQ items 56 and 57 were removed from further analysis. In addition, PQ14, fell in the "mediocre" range was removed on account of the disparity between it and the MSA levels of the remaining variables. Analyses were rerun with these variables removed. All resulting MSA values were in the "middling" range or above.

Three additional variables were excluded from the following analysis. PQ 48 ("I have crises where I call my therapist") was removed on theoretical grounds previously described. The remaining items, PQ41 ("I have trouble falling asleep") and PQ58 ("I see/hear things that other people do not see/hear") were removed from the analysis due to low communality values, .461 and .341, as recommended by Hair et al. (1995). Thus, a dataset comprised of 54 items was submitted for final analysis.

Prior to final analysis, a check on the quality of the correlation matrix for factoring was conducted using the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity. Bartlett's Test of Sphericity provides a description of the adequacy of the data for the purpose of factor analysis. In other words, it is the probability that the observed data are correlated to a significant degree (Hair et al., 1995). The KMO = .912, which indicates that the correlation matrix is in excellent condition for factoring, and Bartlett's Test of Sphericity was significant, χ^2 (1431) = 7856.274, p = .0001.

As warned by Hair et al. (1995), PCA in excess of fifty variables often yields a large number of factors, the latter of which are regularly poorly defined by one or two variables. Indeed, following the latent root criterion, thirteen factors were extracted with eigenvalues greater than one. Because the latter factors were poorly defined, the

determination was made to limit the number of factors to be interpreted by extracting the number needed to account for approximately 60% of the variance, as discussed in Hair et al. (1995). The average communality value was .609. Communalities are presented in Table 6.

Table 6: PQ PCA Communalities

Question	Initial	Extraction	Question	Initial	Extraction
PQ01	1.000	.651	PQ09	1.000	.681
PQ02	1.000	.533	PQ10	1.000	.658
PQ03	1.000	.605	PQ11	1.000	.675
PQ04	1.000	.644	PQ12	1.000	.538
PQ05	1.000	.570	PQ13	1.000	.632
PQ06	1.000	.617	PQ15	1.000	.692
PQ07	1.000	.557	PQ16	1.000	.507
PQ08	1.000	.560	PQ17	1.000	.621
PQ18	1.000	.613	PQ38	1.000	.724
PQ19	1.000	.654	PQ39	1.000	.615
PQ20	1.000	.600	PQ40	1.000	.626
PQ21	1.000	.517	PQ42	1.000	.655
PQ22	1.000	.501	PQ43	1.000	.622
PQ23	1.000	.668	PQ44	1.000	.664
PQ24	1.000	.540	PQ45	1.000	.612
PQ25	1.000	.621	PQ46	1.000	.594

Table 6: PQ PCA Communalities (*continued*)

Question	Initial	Extraction	Question	Initial	Extraction
PQ26	1.000	.628	PQ47	1.000	.546
PQ27	1.000	.611	PQ49	1.000	.592
PQ28	1.000	.697	PQ50	1.000	.587
PQ29	1.000	.538	PQ51	1.000	.519
PQ30	1.000	.653	PQ52	1.000	.554
PQ31	1.000	.675	PQ53	1.000	.605
PQ32	1.000	.691	PQ54	1.000	.535
PQ33	1.000	.642	PQ55	1.000	.629
PQ34	1.000	.531	PQ59	1.000	.649
PQ35	1.000	.575	PQ60	1.000	.581
PQ36	1.000	.681			
PQ37	1.000	.597			

Ten factors with eigenvalues greater than 1 were extracted, accounting for 60.9% of the variance. This solution was submitted to Promax rotation, and the rotated factor solution converged in sixteen iterations. The factors were defined and interpreted based on the Factor Pattern Matrix, see Appendix C for individual variable loadings. The Factor Structure Matrix is also presented in Appendix C.

The factors were defined as follows: *I. Depression*. This factor accounted for 29.2% of the variance and is comprised of the following seventeen items: 25, 23, 08, 05, 40, 42, 15, 33, 06, 21, 09, 55, 24, 22, 03, 18, and 60. *II. Interpersonal Satisfaction*. This

factor accounted for 6% of the variance and is comprised of the following eight items: 01, 13, 43, 37, 20, 30, 31, and 12. *III. Somatic Complaints.* This factor accounted for 4.8% of the variance and is comprised of the following six items: 45, 27, 34, 02, 29, and 51. *IV. Anxiety.* This factor accounted for 4.2% of the variance and is comprised of the following five items: 35, 59, 36, 10, and 47. *V. Primary Role Dysfunction*. This factor accounted for 3.4% of the variance and is comprised of the following five items: 38, 28, 04, 50, and 39. *VI. Alcohol Misuse*. This factor accounted for 3.2% of the variance and is comprised of the following three items: 11, 32, and 26. *VII. Love Relationship Distress*. This factor accounted for 2.9% of the variance and is comprised of the following three items: 17, 07, and 16. *VIII. Dangerousness.* This factor accounted for 2.8% of the variance and is comprised of the following three items: 49, 44, and 46. *IX. Agitation*. This factor accounted for 2.3% of the variance and is comprised of the following two items: 19 and 54. *X. Psychomotor Disturbance*. This factor accounted for 2.1% of the variance and is comprised of the following two items: 52 and 53.

Validation of this empirically derived factor structure through CFA was not possible due to sample size limitations. That is, the sample size did not permit the retention of an independent subsample of data for such an analysis.

For this study, data were to be collected across multiple administrations of the PQ; however, immediately following collection of the data analyzed above, the outpatient facility where they were collected closed. Thus, to permit fulfillment of the original aims of this study, a supplementary dataset was obtained from OQ Systems, Inc., the company that licenses the rights to the OQ-45.2.

3. RELIABILITY & VALIDITY OF THE OUTCOME QUESTIONNAIRE

3.1. Method

3.1.1. Participants

A large sample of data (N = 17,978) was obtained through OQ Systems, Inc., the company that licenses the outcomes management system. Participants were individuals who were receiving outpatient mental health services between November 1, 1998 and February 28, 2002. The database contains repeated measurements for some patients. Descriptive data pertaining to the number of OQ-45.2 administrations are presented in Table 7. However, no other demographic and descriptive data were available on the participants in this study. Random samples were drawn from the database in order to supply the sample data for analyses presented below.

For each analysis, a sample of 450 cases was drawn randomly from the 13,502 baseline administrations of the OQ-45.2 using SPSS Select Cases. This sample size was selected on the basis of available guidelines for sample sizes in factor analytic and SEM analyses, here 10 observations per variable (Kline, 1998).

Table 7: OQ45.2 Administrations

#	N (%)	
1	13,502	(75%)
2	2,483	(14%)
3	801	(4%)
4	398	(2%)

Table 7: OQ45.2 Administrations (continued)

#	N	(%)
5	251	(1%)
6	156	(.9%)
7	106	(.6%)
=8	281	(.250

3.1.2. Measure

The measure being investigated in Study 2 is the Outcome Questionnaire (OQ-45.2; See Appendix C). The OQ-45.2 and its psychometric properties is described thoroughly in Chapter 1. The measure is available in multiple administration methods: Paper-and-pencil, telephone (IVR), internet, computer, and handheld device. Associated with each data collection method is a data entry method, including: manual, scanning, direct, and Hotsync entry. The questionnaire instructions and response choices are identical to those described in Study 1.

3.1.3. Design & Procedures

Specific data collection procedures are not available for this study as a result of the multiple sources of data contributing to the complete data set. All data were collected in outpatient clinical settings throughout the United States using one of the available administration and data entry methods just described.

3.2. Results (Random Sample 1)

3.2.1 Descriptive Statistics

Questionnaire completion and data integrity. The dataset just described contains no missing data points. Each patient's questionnaire was entered by OQ Systems, Inc. directly into an electronic database and scored automatically. Greater than five missing responses resulted in an invalid profile, and the missing items on questionnaires with five or fewer missing were estimated based on the patients' responses to the remaining questions (Maureen Hart, personal communication, September 6, 2002).

3.2.2. Internal Consistency

Cronbach's Coefficient Alpha was computed using SPSS Reliability Analysis to assess the internal consistency of the OQ-45.2. Results for the three theoretically derived subscales of the OQ-45.2 (SD, IR, and SR) are presented in Table 8.

Table 8: OQ-45.2 Total and Subscale Internal Consistency Values

OQ-45.2 Scale	Internal consistency (N = 13,502)
Symptoms distress	.932
Interpersonal relationships	.802
Social role	.740
OQ total	.941

Additionally, Cronbach's Coefficient Alpha was computed for additional administrations of the OQ-45.2. Results are presented in Table 9 for the theoretical subscales as well as the OQ Total score.

Table 9: OQ-45.2 2nd to 5th Assessment Internal Consistency

OQ-45.2 Scale	Internal consistency
2nd Assessment (n = 2483)	
Symptoms Distress	.939
Interpersonal Relationships	.808
Social Role	.762
OQ Total	.947
3rd Assessment (n = 801)	
Symptoms Distress	.946
Interpersonal Relationships	.812
Social Role	.781
OQ Total	.951
4th Assessment (n = 398)	
Symptoms Distress	.948
Interpersonal Relationships	.808
Social Role	.775
OQ Total	.952
5th Assessment (n = 251)	
Symptoms Distress	.950

Table 9: OQ-45.2 2nd to 5th Assessment Internal Consistency (*continued*)

5th Assessment (n = 251)	
Interpersonal Relationships	.833
Social Role	.793
OQ Total	.957

3.2.3. Test-Retest Reliability

Test-retest reliability was examined first in cases with a one week interval between OQ-45.2 administrations. Test-retest reliability was calculated as the bivariate correlation coefficient between the OQ-45.2 subscale and total scores for the first and second administrations of the questionnaire. Results are presented in Table 10 below.

Table 10: OQ-45.2 Test-Retest Reliability

1 Week Interval (N = 1281)	
OQ-45.2 Scale	Pearson r
Symptoms Distress	.800, p = .01
Interpersonal Relationships	.764, p = .01
Social Role	.725, p = .01
OQ Total	.786, p = .01
2 Week Interval (N = 470)	
Symptoms Distress	.747, p = .01
Interpersonal Relationships	.744, p = .01

Table 10: OQ-45.2 Test-Retest Reliability (*continued*)

OQ-45.2 Scale	Pearson r
Social Role	.675, p = .01
OQ Total	.739, p = .01
3 Week Interval (N = 296)	
Symptoms distress	.647, p = .01
Interpersonal relationships	.625, p = .01
Social role	.580, p = .01
OQ total	.618, p = .01
4 Week Interval (N = 188)	
Symptoms distress	.843, p = .01
Interpersonal Relationships	.709, p = .01
Social role	.771, p = .01
OQ total	.821, p = .01
5 Week Interval (N = 133)	
Symptoms Distress	.726, p = .01
Interpersonal Relationships	.745, p = .01
Social role	.660, p = .01
OQ total	.726, p = .01

3.2.4. OQ-45.2 Three Factor CFA

Conceptual model. The conceptual model tested in this analysis was identical to that previously described for the PQ, though only utilizing the original 45 items of the

OQ-45.2. The SD subscale is comprised of the following items: PQ02, PQ03, PQ05, PQ06, PQ08, PQ09, PQ10, PQ11, PQ13, PQ15, PQ22, PQ23, PQ24, PQ25, PQ27, PQ29, PQ31, PQ33, PQ34, PQ35, PQ36, PQ40, PQ41, PQ42, and PQ45. The IR subscale is comprised of the following items: PQ01, PQ07, PQ16, PQ17, PQ18, PQ19, PQ20, PQ26, PQ30, PQ37, and PQ43. Likewise, the SR subscale is comprised of the following items: PQ04, PQ12, PQ14, PQ21, PQ28, PQ32, PQ38, PQ39, and PQ44. See the conceptual model description of the PQ for further details on the model tested below.

Statistical model. The statistical model was likewise similar to that presented above, with the absence of the additional items. In this model, there were a total of 45 observed indicator variables (OQ01-OQ45), 48 unobserved variables (three latent constructs and 45 error terms associated with each OQ-45.2 item), for a total of 93 variables in the model. This yielded a total of 45 free parameters and 63 fixed parameters. Thus, the χ^2 test for the overall fit of the model was based on 1,035 sample moments, 93 parameters to be estimated, and 942 degrees of freedom.

Matrix to be analyzed. Analyses below were based on the covariance matrices of a random sample of 450 baseline OQ-45.2 administrations. Due to the large number of variables in the present analysis, the bivariate correlation matrix and standard deviations have been omitted and are available upon request.

Distribution of data. Univariate estimates of skewness and kurtosis were obtained through SPSS Descriptives and are summarized below. The same guidelines as previously described were used to determine significance. Three variables were found to significantly depart from univariate normality in terms of skewness and kurtosis. The three variables are those pertaining to alcohol misuse (PQ11, PQ26, and PQ32). These

variables were transformed so as to decrease their departures from univariate normality. See Table 11 for the original and transformed skewness and kurtosis values.

With respect to multivariate normality, Mardia's measure of multivariate kurtosis was 316.498, with a critical ratio of 51.615. This indicates that the multivariate distribution of the OQ sample data is significantly non-normal.

Table 11: OQ-45.2 Items with Non-Normal Distributions

OQ Item	Skewness	Kurtosis
OQ11	3.50	14.06
OQ26	3.22	10.97
OQ32	4.58	24.02
OQ11 Transformed	-2.52	4.67
OQ26 Transformed	-2.11	3.18
OQ32 Transformed	-3.28	9.20

Model estimation and fit criteria. Model parameters were estimated through the Maximum Likelihood (ML) method. The same fit indexes as above were used to determine the degree of fit between the observed data and the theoretically specified model.

The regression coefficient for one variable, OQ14 ("I work/study too much") was non-significant; accordingly, this variable was removed from subsequent analyses. The model parameters are presented in Appendix D. Visual inspection reveals appropriate values for the standardized regression coefficients and variances.

Because of the excessively kurtotic nature of the data, the bootstrapping procedure was employed. See Study 1 for a description of and rationale for this procedure. The Bootstrapping parameter estimates obtained through this procedure are omitted due to space considerations and available by request.

Evaluation of model fit. The overall test of model fit indicated that the theoretically derived three factor model tested here was not plausible based on the observed data, $\chi^2(942, N = 450) = 3072.740$, $p = .0001$, GFI = .732, TLI = .760, and CFI = .772. The χ^2 was also evaluated according to the Bollen-Stine bootstrap method and, based on the empirically derived normal sampling distribution, was still significant, $p = .002$.

3.2.5. Summary

The focus next turned to respecification of the original model such that the sample data more closely correspond to the specified model. It must be noted explicitly that, at this stage, examination of the factor structure of the OQ-45.2 changes from a confirmatory approach to an exploratory approach (Byrne, 2001). Principle components analysis was used for this purpose, and, once the empirically derived factor structure was obtained, the resulting factor structure was validated through CFA procedures using a separate random sample of baseline data.

3.3. Exploratory Derivation of the OQ-45.2 Factor Structure

A Principle Components Analysis (PCA) with Promax rotation was performed on the same sample of data using SPSS Factor Analysis. The rationale for selecting these procedures was previously described. The bivariate correlation matrix was first inspected to establish the degree of variable inter-correlation. Approximately 35% of the observed

correlations were at least .30 in magnitude (i.e., 346 of 990). The average correlation among all pairs was .24. Of the residuals, 1.4% were greater in absolute value than .10. Of these, none were in excess of .20 or .30 in magnitude. Thus, with the effects of the other variables removed, only a small portion of variables retained a significant bivariate relationship. This is a desirable condition for factor analysis.

Univariate MSA values were examined prior to interpretation of the factor solution. One variable, OQ14 ("I work/study too much"), fell in the "unacceptable" range with a value of .459, and was subsequently removed from further analysis. Also of note, the three alcohol misuse related items fell in the "miserable" or "mediocre" range, with values of .616, .594, and .621, for OQ11 ("After heavy drinking, I need a drink the next morning to get going"), OQ26 ("I feel annoyed by people who criticize my drinking"), and OQ32 ("I have trouble at work/school because of drinking or drug use"), respectively. These variables were retained for analyses on the basis of the previous PCA which suggested that they form a unique factor. Finally, OQ41 ("I have trouble falling asleep or staying asleep") and OQ12 ("I find my work/school satisfying") were removed on the basis of low communalities, .389 and .491, respectively. Thus, a dataset comprised of 42 items was submitted for final analysis.

Prior to final analysis, a check on the quality of the correlation matrix for factoring was conducted using the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity. The KMO = .927, which indicates that the correlation matrix is in excellent condition for factoring, and Bartlett's Test of Sphericity was significant, χ^2 (861) = 8716.109, p = .0001.

The latent root criterion was applied to extract nine factors with eigenvalues greater than one (Gorsuch, 1988), accounting for approximately 62% of the variance. The average communality value was .617. Communalities are presented in Table 12 below.

Table 12: OQ-45.2 PCA Communalities

Question	Initial	Extraction	Question	Initial	Extraction
OQ01	1.000	.646	OQ23	1.000	.632
OQ02	1.000	.646	OQ24	1.000	.646
OQ03	1.000	.584	OQ25	1.000	.551
OQ04	1.000	.609	OQ26	1.000	.734
OQ05	1.000	.526	OQ27	1.000	.546
OQ06	1.000	.664	OQ28	1.000	.700
OQ07	1.000	.669	OQ29	1.000	.531
OQ08	1.000	.647	OQ30	1.000	.577
OQ09	1.000	.616	OQ31	1.000	.694
OQ10	1.000	.638	OQ32	1.000	.683
OQ11	1.000	.593	OQ33	1.000	.560
OQ13	1.000	.655	OQ34	1.000	.570
OQ15	1.000	.719	OQ35	1.000	.518
OQ16	1.000	.510	OQ36	1.000	.549
OQ17	1.000	.581	OQ37	1.000	.600
OQ18	1.000	.593	OQ38	1.000	.724
OQ19	1.000	.640	OQ39	1.000	.715

Table 12: OQ-45.2 PCA Communalities (*continued*)

Question	Initial	Extraction			
OQ20	1.000	.675	OQ40	1.000	.582
OQ21	1.000	.516	OQ42	1.000	.647
OQ22	1.000	.559	OQ43	1.000	.589
OQ44	1.000	.706			
OQ45	1.000	.558			

This solution was submitted to Promax rotation, and the nine factor rotated solution converged in thirteen iterations. The factors were defined and interpreted based on the Factor Pattern Matrix, see Appendix E1 for individual variable loadings. The Factor Structure Matrix is also presented in Appendix E2. The factors were defined as follows: *I. Depression*. This factor accounted for 29.2% of the variance and is comprised of the following eleven items: OQ24, OQ13, OQ31, OQ21, OQ43, OQ03, OQ20, OQ23, OQ15, and OQ42. II. Anxiety. This factor accounted for 7.0% of the variance and is comprised of the following seven items: OQ10, OQ36, OQ33, OQ35, OQ29, OQ25, and OQ05. *III. Primary Role Dysfunction*. This factor accounted for 5.4% of the variance and is comprised of the following four items: OQ38, OQ28, OQ04, and OQ39. *IV. Love Relationship Distress*. This factor accounted for 4.6% of the variance and is comprised of the following four items: OQ17, QO37, OQ07, and OQ18. *V. Somatic Complaints.* This factor accounted for 3.7% of the variance and is comprised of the following five items: OQ02, OQ34, OQ45, OQ27, OQ09, and OQ22. *VI. Alcohol Misuse*. This factor accounted for 3.3% of the variance and is comprised of the following three items: OQ26,

OQ32, and OQ11. *VII. Dangerousness.* This factor accounted for 3.0% of the variance and is comprised of the following three items: OQ44, OQ08, and OQ40. *VIII. Conflict*. This factor accounted for 2.7% of the variance and is comprised of the following three items: OQ19, OQ16, and OQ06. *IX. Interpersonal Relationship Difficulty.* This factor accounted for 2.5% of the variance and is comprised of the following two items: OQ01 and OQ30.

The empirically derived nine factor structure just described was next submitted to CFA for the purpose of validation using a new sample of 450 cases. The results are described below.

3.4. Validation of Empirically Derived Factor Structure (Random Sample 2)

Conceptual model. The conceptual model tested is identical to that of the OQ-45.2 empirically derived nine factor structure described in the previous section. That is, the model is specified such that the variables loading on the nine latent constructs are those listed as defining the above factors.

Statistical model. In this model, there are a total of 42 observed indicator variables (OQ01-OQ45 with OQ12, OQ14, and OQ41 removed), 51 unobserved variables (9 latent constructs and 42 error terms associated with each OQ-45.2 item), for a total of 93 variables in the model. This yields a total of 42 free parameters and 51 fixed parameters. Thus, the χ^2 test for the overall fit of the model is based on 903 sample moments, 120 parameters to be estimated, and 783 degrees of freedom.

Matrix to be analyzed. Analyses below are based on the covariance matrices of a new random sample of 450 baseline OQ-45.2 administrations. The bivariate correlation matrix and standard deviations for all OQ-45.2 items are available upon request.

Distribution of data. Univariate estimates of skewness and kurtosis were obtained through SPSS Descriptives and are summarized below. The guidelines discussed above were applied to determine significance. Similarly to above, the three items relating to alcohol misuse were found to depart significantly from univariate normality (PQ11, PQ26, and PQ32). Table 13 presents the univariate skewness and kurtosis values for these items.

Multivariate kurtosis was assessed by estimates provided by AMOS based on Mardia's test of multivariate normality (Byrne, 2001; Mardia, 1970). Mardia's measure of multivariate kurtosis was 295.912, with a critical ratio of 51.626. Thus, the sample data depart significantly from multivariate normality.

Table 13: Non-Normally Distributed OQ-45.2 Validation Sample Items

OQ Item	Skewness	Kurtosis
OQ11	4.19	21.99
OQ26	3.17	10.00
OQ32	3.99	16.47
OQ11 Transformed	-2.59	5.15
OQ26 Transformed	-2.39	4.02
OQ32 Transformed	-3.20	8.64

Model estimation and fit criteria. Model parameters were estimated through the Maximum Likelihood (ML) method. All regression coefficients were significant; and accordingly, all variables are retained in this analysis. Visual inspection of the

standardized regression coefficients and variances presented in Appendix F revealed than no values fell outside of the acceptable range.

Because of the excessively kurtotic nature of the data, the bootstrapping procedure was employed. See Analysis 1 for a description of this procedure. Appendix F contains the parameter estimates obtained through this procedure.

Evaluation of model fit. The overall test of model fit indicates that the empirically derived nine factor model obtained through PCA is not plausible based on an independent sample of data, $\chi^2(824, N = 450) = 2332.413$, $p = .0001$, GFI = .787, TLI = .819, and CFI = .835. The χ^2 was also evaluated according to the Bollen-Stine bootstrap method, and when using an empirically derived normally distributed sampling distribution was still significant, $p = .002$. However, the fit of this model is a statistically significantly improved solution to the three-factor model tested above, $?\chi^2(118) = 740.327$, $p < .0001$.

4. OUTCOME QUESTIONNAIRE FACTOR STRUCTURE STABILITY

This study examined the degree of consistency between the factor structure of the OQ-45.2 at the baseline administration and a later administration. This was done so as to determine whether the factor structure of the questionnaire remains stable despite elapsed time and receipt of treatment intervention or whether the effect of time and intervention is such that the underlying constructs measured by the OQ-45.2 are variable.

4.1. Method

4.1.1. Sample

The fourth administration of the OQ-45.2 was selected to provide a comparison group for addressing this question There were 398 cases with at least four administrations of the OQ-45.2. Of these, cases were selected in which there was no more than a twelve week interval between the baseline administration and the fourth administration of the questionnaire. This yielded 341 cases, or 85.4% of the original sample. The remaining cases were excluded from analysis. The average time interval between the baseline and fourth administrations was, M = 5.3 weeks (SD = 2.6).

4.2. Results

4.2.1. OQ-45.2 Internal Consistency

Cronbach's Coefficient Alpha was computed using SPSS Reliability Analysis to assess the internal consistency of the OQ-45.2 at the first and fourth administrations. Results for the three theoretically derived subscales of the OQ-45.2 (SD, IR, and SR) are presented in Table 14.

Table14: Internal Consistency Values for the OQ-45.2 and Subscale Scores

Scale	Internal consistency
1st Administration (N = 450)	
Symptoms distress	.934
Interpersonal relationships	.805
Social role	.718
OQ total	.939
4th Administration (N = 341)	
Symptoms distress	.943
Interpersonal relationships	.811
Social role	.758
OQ total	.948

4.2.2. Comparison of Baseline and Fourth Administration Factor Structures

Distribution of data. Three variables were found to have significant univariate departures from normality in terms of skewness and kurtosis for each sample. These variables are the three pertaining to alcohol misuse (PQ11, PQ26, and PQ32). See Table 15 for the actual skewness and kurtosis values for these variables at each timepoint.

Mardia's measure of multivariate kurtosis for the two groups combined was 338.676, with a critical ratio of 57.744. This indicates that the extent of multivariate non-normality present in the OQ sample is significant. The transformed variables reduced Mardia's measure of multivariate kurtosis to 301.563 with a critical ratio of 51.416. Thus, despite transformation, the multivariate distribution remains significantly non-normal.

Table 15: Non-Normally Distributed OQ-45.2 Items

OQ Item	Skewness	Kurtosis
1st Administration		
OQ11	3.98	20.35
OQ26	3.50	13.92
OQ32	5.37	32.55
OQ11 Transformed	- 2.57	4.99
OQ26 Transformed	- 2.29	3.61
OQ32 Transformed	- 3.85	13.54
4th Administration		
OQ11	5.57	36.07
OQ26	4.68	22.78
OQ32	4.73	26.47
OQ11 Transformed	- 4.19	16.27
OQ26 Transformed	- 3.38	10.08
OQ32 Transformed	- 3.51	10.79

Testing for Invariance. The procedure utilized below allows for comparisons to be made with respect to the degree of fit between a specified factor structure for two groups. In this case, the first group is comprised of patients who completed baseline administrations of the OQ-45.2 and the second group of patients who completed the fourth administration.

The empirically derived nine factor structure detailed above was imposed on the fourth administration data to determine the corresponding degree of fit. This was accomplished through several steps. First, the models were compared with no equality constraints imposed on the parameters, other than those required for the purpose of model identification. As recommended by Byrne (2001), multi-group model fit was assessed through χ^2, CFI, and RMSEA. While the first two fit indices were detailed in Study 1, the Root Mean Square Error of Approximation (RMSEA) has not been utilized to this point. RMSEA provides an indication of the degree of fit between a hypothesized model and observed data where values less than .06 represent good fit and values from .08-.10 represent mediocre fit (Byrne). RMSEA has an associated probability value, which should be in excess of .50, as well as confidence intervals.

In this model the chi-square test was significant, χ^2 (1648) = 4655.004, $p < .0001$; CFI = .821; and RMSEA = .048, p = .973. This provides the baseline model, the model to which subsequent models were compared.

The second step involved placing equality constraints on all regression coefficients, latent construct variances, and factor covariances. In other words, these parameters were statistically forced to be equivalent across the two groups, baseline and fourth administration. Model fit was reassessed. The goodness of fit statistics were, χ^2 (1726) = 4986.185, $p > .05$, CFI = .799, and RMSEA = .049, p = .866. The chi-square difference (i.e., the difference between this χ^2 value and that of the baseline model) and the associated degrees of freedom difference were tested for significance, $\Delta\chi^2$ (78) = 331.181, $p < .05$. This indicates that, with all model parameters constrained to equality, the two models are not invariant.

Third, only the factor loading constraints were retained (i.e., required to be statistically equivalent across the two groups). Results revealed that $\Delta\chi^2$ (34) = 267.411, $p > .05$, CFI = .800, RMSEA = .049, p = .724. This indicates that with the factor loadings constrained to equivalency, the two models were not invariant.

Table 16: OQ-45.2 Model Invariance Goodness of Fit Statistics

Model Description	Comparative Model	χ^2	Df	$\Delta\chi^2$	Δdf	Significance
Hypothesized Model	Model 1	4655.004	1648	-	-	-
All Parameters Constrained	Model 1	4986.185	1726	331.181	78	$p < .05$
All Factor Loadings Constrained	Model 1	4922.415	1682	267.411	34	$p < .05$
Factor 1 Constrained	Model 1	4899.829	1658	244.825	10	$p < .05$
All Construct Variances Constrained	Model 1	4888.765	1657	233.761	9	$p < .05$
Depression Constrained	Model 1	4882.697	1649	227.693	1	$p < .05$

Table 16: OQ-45.2 Model Invariance Goodness of Fit Statistics (*continued*)

Model Description	Comparative Model	χ^2	Df	$\Delta\chi^2$	Δdf	Significance
Anxiety Constrained	Model 1	4882.622	1649	227.618	1	$p < .05$
Role Constrained	Model 1	4883.731	1649	228.727	1	$p < .05$
Somatic Constrained	Model 1	4882.335	1649	227.331	1	$p < .05$
Love Constrained	Model 1	4884.653	1649	229.649	1	$p < .05$
Alcohol Constrained	Model 1	4882.416	1649	227.412	1	$p < .05$
Danger Constrained	Model 1	4883.562	1649	228.558	1	$p < .05$
Relationships Constrained	Model 1	4882.303	1649	227.229	1	$p < .05$
Conflict Constrained	Model 1	4883.101	1649	228.097	1	$p < .05$

The next task was to determine the equivalency of the statistical model, which involved constraining to equivalency the variances of each of the nine factors. Results indicated that the nine factors were not invariant, $\Delta\chi^2$ (9) = 233.761, p = .05, CFI = .801,

RMSEA = .050, p = .610. Following this, the variance of each of the nine factors was constrained sequentially. If the variance of a given factor was determined to be invariant across the two groups, this constraint was retained. Otherwise, the constraint was freed and freely estimated. Results revealed than none of the variances of the nine factors was invariant between the groups.

5. DISCUSSION & CONCLUSIONS

5.1. Discussion

Utilizing outpatient mental health samples, the two main goals of this investigation were to examine the psychometric properties of the Progress Questionnaire and to provide further evidence of the reliability and validity of the Outcome Questionnaire. In addition, an overarching aim of this investigation was to determine the usability of the PQ for the purpose of monitoring patient progress in treatment and conducting patient focused research in outpatient clinical settings. These goals were accomplished through three studies, which are summarized below.

Studies 1 and 2 provided an empirical test of the hypothesis that the PQ and OQ-45.2 measure three latent constructs that are considered to be important in the assessment of treatment change: Symptom Distress, Interpersonal Relations, and Social Role. The results of confirmatory tests of this hypothesis in Studies 1 and 2 failed to support the presence of this three factor structure. Indeed, the results of the exploratory factor analytic procedures revealed that the PQ and OQ-45.2 are comprised of ten and nine correlated factors, respectively. In addition, each study examined the reliability of these measures and found that the PQ and the OQ-45.2 each possess a desirable degree of internal consistency. Study 3 failed to support the hypothesis that the factor structure of the OQ-45.2 is invariant across repeated administrations, indicating that the domains assessed by the OQ-45.2 vary significantly between the baseline and fourth administration. Finally, in Study 1, the PQ was found to be readily implemented in an outpatient clinical setting. These findings, as well as the challenges and considerations associated with the clinical use of the PQ are discussed further below.

Because the three studies just summarized were highly consistent in their objectives and outcomes, the major findings and conclusions of each are discussed collectively below. The discussion now turns to examination of these findings. The discussion that follows includes an evaluation and explanation of the results, explication of the contribution of these findings, consideration of the strengths and limitations of each study, and recommendations regarding future clinical and research applications of the OQ-45.2.

5.1.1. Major Findings

The results of Study 1 establish the psychometric properties of the PQ, and Study 2 yields further evidence for the operating characteristics of the OQ-45.2 The findings in each study are mixed. The theoretically derived three factor structure that each measure is purported to possess was not supported. In fact, the fit of the PQ and OQ-45.2 sample data to this factor structure was extremely poor, and the addition of the fifteen new items to the PQ in Study 1 did not improve this fit.

This finding is at least partially consistent with the limited existing empirical evidence for the factor structure of the OQ-45.2. In a published confirmatory test of the theoretically derived factor definition, Mueller et al. (1998) found only weak support for the existence of the three latent constructs. However, the present studies depart from this finding in that the fit of the three factor solution obtained here was substantially worse than that reported by Mueller et al. Overall, then, some question exists as to the validity of the theoretical factor structure of the OQ-45.2 based on previous findings and the findings of the present studies. These results suggest that the latent constructs assessed by

the PQ and the OQ-45.2 are not defined by the groupings of items comprising the SD, IR, and SR subscales in heterogeneous clinical samples.

The PQ and OQ-45.2 were found to possess desirable reliability properties. The internal consistency estimates obtained in Studies 1 and 2 are consistent with those reported in the existing literature on the OQ-45.2 (e.g., Lambert & Finch, 1999). On each measure, the Total Score and SD subscales possess excellent internal consistency, and the IR and SR subscales possess acceptable internal consistency. For the OQ-45.2, these estimates remained stable across repeated administrations of the questionnaire and with time intervals of various lengths between administrations, further solidifying this property of the measure. Similarly, the estimates of test-retest reliability for the OQ-45.2 are consistent with those reported by Lambert and Finch, with the strongest relationships between administrations occurring between the Total Score and SD subscale at the two timepoints.

These findings provide further evidence of the excellent reliability of the OQ-45.2 as a global measure of symptomatology and functioning and the relative weakness of the IR and SR subscales in comparison to the Total Score and SD subscale. This, combined with previous research findings, suggests that the IR and SR scales are less well defined and possess greater heterogeneity of item content that the SD subscale and Total Score. Also significant are the test-retest reliability findings. Though consistent with previously reported estimates test-retest reliability obtained in an all student sample (Lambert et al., 1996), these estimates provide evidence of the temporal stability of the OQ-45.2 in an all clinical outpatient sample.

Because of the excessively poor fit of the three factor structure discussed above, exploratory factor analytic procedures were undertaken in Studies 1 and 2 so as to provide clues, derived from a strictly empirically based definition of the factor structure, as to the constructs that are assessed by these questionnaires. These results provide evidence of a more complex factor structure than that originally proposed for the PQ and OQ-45.2, with ten or nine correlated factors on the PQ and OQ-45.2, respectively. This suggests that both the PQ and OQ-45.2 are comprised of a larger number of smaller, more specific factors than intended in their original design. These analyses produced desirable relationships among the questionnaire items, indicating that the PQ and OQ-45.2 are valid measures.

Due to sample size limitations in Study 1, the empirically derived ten factor structure was not validated using an independent sample of data, though this procedure was utilized for the OQ-45.2 in Study 2. Although validation of the nine factor structure revealed that the model was implausible with an independent sample of data, the model did provide significantly improved fit to the three factor model. This provides further evidence against the theoretically derived three factor structure and evidence in support of nine factors comprising the OQ-45.2.

The results of Study 3 indicate that the OQ-45.2 factor structure resulting from exploratory procedures using a sample of baseline data does not hold when tested at the fourth administration of the questionnaire. One possible explanation for this is that the factor structure of the OQ-45.2 varies across time and as a result of treatment intervention. In other words, the underlying constructs assessed by this questionnaire may not be stable across time and with the amelioration of symptoms and improvement in

functioning expected as the result of treatment intervention. In addition, the relatively poor fit obtained in Study 2 may be responsible for the poor fit obtained at a later time. That is, the factor structure tested at each time-point fit poorly but may have fit poorly in different ways. This is an interesting consideration that deserves continued attention once the constructs assessed by the OQ-45.2 are more accurately defined. Given the results of Studies 1, 2, and 3 just discussed, it is important to address potential explanations for this poor fit. This is undertaken below.

5.1.2. Factors Contributing To Fit

Specifying the underlying factor structure of the PQ and OQ-45.2 is an important task for more than purely theoretical reasons. These questionnaires are designed to inform actual clinical practice by providing feedback to the clinician about treatment progress. Clinically, then, the instruments' subscales are used to inform treatment planning and clinical decision making. Likewise, in the context of managed care, these questionnaires may be utilized to inform decisions regarding the authorization of treatment sessions. As a result, it is critical to be confident about the constructs assessed by these questionnaires. The findings outlined above call into question the validity of the theoretical three factor structure of the PQ and OQ-45.2, suggesting that they are comprised of a larger number of factors. Following is a discussion of the possible explanations for these findings as well as methods of and considerations in modifying the questionnaires so as to improve the definition of the factors.

There are several possible factors that contribute to the poor fit obtained in the analyses discussed above. The most obvious explanation is that of factor

misspecification, or the inaccurate definition of the PQ and OQ-45.2 as measures of three latent constructs, SD, IR, and SR. Evidence of this position was reviewed above.

In addition to factor misspecification, there are other potential influences on the poor fit of the theoretical factor structure. One possible source of misfit in factor analytic studies is the existence of distinct subgroups within the sample whose responses vary systematically from those of other groups (Hair et al., 1995). Considering these subgroups together as one group, then, serves to obscure the underlying factor structure.

In this study, there are two main categories of variables around which distinct subgroups potentially exist, diagnostic and demographic variables. First, in contrast to previous studies on the OQ where approximately 40% of the sample was non-clinical (i.e., community and student normals), the samples analyzed here were comprised entirely of individuals receiving psychotherapy and or medication management services. As a result, the possibility exists that distinct diagnostic subgroups exist within the samples of data (e.g., depressive disorders and psychotic disorders). However, the data and sample size requirements necessary to test this hypothesis were not available in Study 1 or 2. In Study 1, sample size limitations and the limited availability of detailed diagnostic data were prohibitive. Likewise, in Study 2, diagnostic data were not available. Similarly, the second potential source of differential subgroup responding to the PQ and OQ-45.2 items is that of systematic response differences on key demographic variables such as age, gender, ethnicity, socio-economic status, etc.

Modification of the questionnaire's content is another possible method of improving the factor definition of the OQ-45.2, and the results of Studies 1 and 2 identify some consistently problematic items. Two items were consistently found to perform

poorly based on statistical criteria in the PQ and OQ-45.2 samples. Questions 14 ("I work/study too much") and 41 ("I have trouble falling asleep or staying asleep") were poorly related to the other variables in the analyses. Mueller et al. (1998) also found Item 14 to perform poorly. Thus, future studies and applications of these questionnaires should carefully examine these items for evidence of their utility and/or consider them for removal. In addition, recommendations for improving the performance of these items are detailed below.

The three items related to alcohol misuse, 11, 26, and 32, were also found to be problematic in this investigation, as they were significantly non-normally distributed in each study (with the exception of Item 26 in Study 1) and poorly related to the other questionnaire items. However, these items consistently formed a unique factor in the exploratory analyses. Examination of the patients' responses revealed that these items were infrequently endorsed, which raises the issue of the appropriateness of alcohol related questions on a questionnaire of symptomatology and functioning designed for an outpatient clinical population, as well as the willingness of patients to openly endorse alcohol misuse in this setting.

The alcohol use items were clearly problematic on statistical grounds; however, with respect to clinical utility, alcohol use is an important domain to assess. Empirical criteria, therefore, should not be the sole determinant of the questionnaire makeup with regard to low baserate or highly clinically salient behaviors. Future research, and research examining the use of the PQ and OQ-45.2 in a variety of clinical settings and populations, should continue to evaluate the functioning and utility of these items, both statistically and clinically.

5.1.3. Evaluation of Clinical Implementation of the PQ

A more general aim of this investigation was to examine the usability of the PQ in a patient focused research context. As previously stated, a fundamental goal of patient focused research endeavors is the measurement and evaluation of patient functioning with minimal interruption to "practice as usual," referring to the minimal additional placement of burden on the provider, office staff, and patient.

Study 1 provides useful information on the implementation of the PQ. The PQ was found to meet the goals just mentioned and was readily implemented into the clinical setting with limited challenges. The office staff owned the responsibility for distributing and collecting the questionnaires from all new patients. However, because this practice was closed shortly after the implementation of the PQ pilot data collection project, the provider feedback aspect of this patient focused research endeavor could not be assessed.

5.1.4. Recommendations for Use of the PQ and OQ-45.2

Examination of the overall utility of the PQ (as an alternative measure to the OQ-45.2) and the OQ-45.2 were additional global aims of this study. The PQ incorporated fifteen new items believed by the MBHO to be of particular importance in measuring symptomatology and patient progress in treatment. However, these additional items did not improve or supplement the reliability and validity of the OQ-45.2.

The content and/or wording of several of the fifteen new items were problematic on theoretical grounds, warranting their removal from the questionnaire and analyses. Four of the six variables found to be statistically or theoretically problematic in Study 1 were new items. More specifically, PQ48, PQ56, and PQ57 presuppose current

involvement in mental health treatment (psychotherapy and/or medication management), which is inappropriate for a measure administered at the initiation of treatment.

The amount of variance accounted for in the PQ and OQ-45.2 exploratory analyses was consistent, with the nine factor structure of the OQ-45.2 accounting for a slightly larger percentage of variance than that of the PQ (62% versus 61%). This indicates that the addition of these items did not serve to more fully explain the underlying constructs. Likewise, the resulting exploratory factor structures were similarly defined, again suggesting that the new items did not improve the definition of the latent constructs. Based on the findings just described, further use of the PQ is not recommended.

Also of particular importance, with respect to the clinical use of these measures, was the amount of missing data present in the completed questionnaires. The patterns of missing data suggest that questionnaire completion rates would likely be improved by making modifications to the questionnaire's instructions and response choices. That is, the missing data revealed that the items most frequently skipped are those that are not applicable to certain patients (e.g., "I feel unhappy in my marriage/significant relationship"), as opposed to those which all patients are able to answer (e.g., "I like myself"). The instructions and response options on the OQ-45.2 should be reworded to plainly reflect that a response choice of zero indicates "Does Not Apply" and/or "Not At All." The response choice of "0=Never" should be changed to "0=Never/Not Applicable."

Questionnaire completion rates would be improved by respecifying the question content of items that are potentially "Not Applicable." For instance, "I feel stressed at

work/school" (OQ04) should be reworded to "I feel stressed in my main role (work, school, home, or volunteer)." Wording that explicitly reflects the variety of activities comprising social role functioning (i.e., work, school, home, volunteer) should be applied to the remaining social role items (OQ12, OQ14, OQ28, OQ32, OQ38, OQ39, and OQ44). By modifying these items, they become applicable to virtually all respondents.

Also related to problematic questions, "I have trouble falling asleep or staying asleep" (OQ-41) was repeatedly found to correlate poorly with the other OQ-45.2 items. However, some concern is noted in removing this item from the questionnaire based on the clinical observation that sleep disturbance is a frequently occurring symptom, relevant target of intervention, and indicator of treatment response when improving. The performance of this item would likely be improved through the modification of its content and the addition of related items. For example, changing OQ41 to "I have difficulty falling asleep" and adding "I wake frequently during the night" and "I wake up early and am unable to fall back to sleep" so as to better define a sleep factor.

On a side note, with respect to missing data, a promising solution to the problem is the increasing availability and application of computer-based assessment procedures. This includes local and web-based assessment packages as well as the use of handheld devices. Computer-based data collection methods ensure complete responses to all items by forcing response choices.

A number of recommendations regarding the future use of the OQ-45.2 can be made based on the results of this investigation. When utilized in a patient-focused research endeavor, the clinician receives a feedback report detailing the patient's scores, endorsement of critical items, and treatment trajectory. Clinicians should continue to use

and be confident in the OQ-45.2 total score as a global indicator of symptomatology and functioning. However, the three theoretical subscales should not be used for the purpose of making treatment related decisions, as support was not found for the existence of these factors. Alternatively, the 9 factor structure detailed in Study 2 should be implemented. There are two main benefits of this approach. First, the nine factor structure enjoys statistically improved fit from the three factor structure, that is, it provides a more accurate definition of the constructs assessed by the OQ-45.2. Second, the nine factors are more specific and inherently more meaningful clinically than the three factor model. Finally, the clinician is encouraged to examine the individual item endorsement of treatment-relevant questions.

The decision to utilize the OQ-45.2 as an outcomes measurement system is largely dependent upon the goals of the clinician, practice, and/or research endeavor. Questionnaires designed for patient-focused research are not best suited for traditional effectiveness and efficacy research designs that allow for the use of multiple assessment instruments and varying degrees of modification to clinical practice during data collection. Likewise, these instruments are not best suited for diagnosis-specific assessment purposes. That is, if the clinical and/or research goal is to provide an accurate measure of the symptomatology associated with a particular diagnosis (e.g., depression), well-established and psychometrically sound diagnosis-specific assessment instruments (e.g., BDI-II) are a more appropriate choice.

However, if the goal of the clinician, practice, and/or research endeavor is ongoing evaluation of patient progress in treatment (in contrast to an episode of data collection for the purpose of answering a specific research question), coupled with the

provision of feedback to the clinician for the purpose of treatment modification, the OQ-45.2 is an excellent choice. More specifically, a substantial body of research exists supporting the internal consistency, test-retest reliability, and, with this investigation, the factor structure of the instrument. In addition, the OQ-45.2 possesses items related to interpersonal functioning and primary role functioning, an improvement upon existing global measures that are limited in scope to the assessment of symptoms. Finally, the OQ-45.2 has the benefit of providing an individual expected treatment response that can be utilized to adjust treatment and evaluated the effectiveness of interventions. This is not meant to suggest that the OQ-45.2 is without limitations. Rather, it enjoys empirical support and is a positive step toward the inclusion of measurement/evaluation as a standard component of routine clinical practice

With respect to future questionnaire modification, a seemingly beneficial feature of many outcome measurement systems similar to the OQ-45.2 is the flexibility to tailor questionnaire item content to specific clinical settings and domains of inquiry. However, as evidenced in Study 1, this practice should be employed extremely judiciously, and any such questionnaire modifications should be made with sound theoretical justification. Because questionnaire modification weakens existing confidence in an instrument's psychometric properties, modifications should be submitted to empirical and statistical scrutiny, as in this investigation, so as to establish the psychometric properties of the new instrument. The discussion next focuses on the strengths and limitations of this investigation.

5.1.5. Strengths and Limitations

A notable strength of this investigation is that the findings detailed above were obtained in an entirely clinical sample. As discussed previously, much of the existing literature on the psychometric properties of the OQ-45.2 were obtained in "normal" populations (e.g., Lambert & Finch, 1999). Thus, Study 2 provides the first known reliability estimates of the measure that are based on an entirely outpatient clinical sample. Likewise, if the PQ is utilized further, the results of Study 1 establish the internal consistency and validity characteristics of the PQ in an outpatient mental health sample, the population for which it is intended. The same is also true of the principle components analysis findings, which represent the first published accounts of exploratory factor analytic procedures describing the underlying constructs of the OQ-45.2.

An additional strength, in contrast to many studies occurring in the social sciences, is that the sample sizes of Studies 1 and 2 were adequate and even desirable. That is, the PQ sample was of sufficient size to rule out sample size limitations as a primary explanation for the poor confirmatory factor analysis findings. Likewise, the extremely large set of data available on the OQ-45.2 permitted the validation of exploratory findings on a separate sample as well as analysis of data across multiple time points. This further solidifies the existing knowledge of the operating characteristics of the OQ-45.2.

An important limitation of these studies is the presence of incomplete data on key demographic and diagnostic variables. Given the poorly fitting factor structure, the availability of such data, with a correspondingly large sample size, would permit more precise investigation of the factor structure of the PQ and OQ-45.2. Unfortunately, the

premature termination of the PQ data collection prevented the attainment of samples of adequate size for such analysis. In addition, the large number of "Diagnosis Deferred" diagnoses in Study 1 was problematic, as was the missing demographic data, as it prevented further evaluation of this hypothesis. These are important areas for consideration in future studies of the OQ-45.2.

5.1.6. Future Directions

Future research should continue to be conducted on the OQ-45.2 in various clinical settings and utilizing diverse clinical populations. As in this investigation, psychometric studies of frequently employed assessment instruments, particularly those providing data used to make decisions affecting patient care, should be routinely conducted with each new use of the questionnaire. A significant body of research is amassing on the OQ-45.2, and these research endeavors should continue. Given the increasingly widespread use of the OQ-45.2, and the proliferation of systems like it, the opportunity exists to convincingly establish this, and other, instrument's psychometric properties while contributing significantly to the understanding of changes that occur through psychotherapy.

A fundamental question that remains involves the underlying factor structure of the OQ-45.2. The results of this study indicate that the OQ-45.2 more accurately measures nine factors despite limited support for the three factor structure. This is a critical issue that should continue to be addressed, taking into consideration the possible explanations discussed above for the problematic fit achieved to date.

Future research should utilize both exploratory and confirmatory factor analytic procedures so as to more precisely specify and define the underlying factor structure of

the OQ-45.2. Likewise, item response theory, and analyses occurring at item level, should be applied so as to better illuminate the functioning and performance of individual items comprising the OQ-45.2. The lack of such investigations in the literature to date is striking provided the number of studies utilizing this questionnaire and the massive amounts of data that have been collected. This may be the result of inadequate investigation of the factor structure or the lack of desirable findings resulting in a limited number of publications. Regardless of the cause, this is an important issue given the current reliance of the OQ-45.2 and the provider feedback on the three subscales. In addition to future factor analytic studies, these studies should consider the effect of diagnostic and demographic subgroups on factor definition. Though the answers to these questions, and a clearly supported factor structure, remain unclear, the performance of the OQ-45.2 as a reliable and valid global measure of symptomatology and functioning is more well-founded. This is an important point, as the findings to date do not suggest that the OQ-45.2 is not valid; rather, they indicated that the subscales are potentially in need of redefinition.

The questions and issues facing the OQ-45.2 just discussed ultimately reduce to the issue of construct validity. Much more confirmatory and exploratory factor analytic work incorporating the recommendations made above needs to be done, but a fundamental question facing the OQ-45.2 is whether the theory underlying its construction (i.e., the constructs important in measuring psychotherapy change are symptom distress, interpersonal relationships, and social role functioning) is in need of respecification.

5.2. Conclusion

In conclusion, this investigation provides further evidence of the psychometric properties of the OQ-45.2 while establishing these properties for the PQ. Such instruments and outcomes measurement systems are being increasingly integrated into routine clinical practice in the context of patient-focused research endeavors. This is an exciting direction for the field, as it seems to provide an ideal marriage of research and clinical practice. The result is an increased understanding of improvements achieved through psychotherapy and medication-based treatment interventions through the clinical implementation of an ongoing treatment response information system. Ideally, this combination of research and clinical practice will continue to proliferate, to the benefit of patients and the current understanding of mental illness and its treatment.

LIST OF REFERENCES

Anastasi, A. (1982). Psychological testing (5th ed.). New York: MacMillan.

Anastasi, A. (1986). Evolving concepts of test validation. Annual Review of Psychology, 37, 1-15.

American Psychological Association (1992). Ethical principles of psychologists and code of conduct. American Psychologist, 47, 1-15.

Berman, W. H., Rosen, C. S., Hurt, S. W., & Kolarz, C. M. (1998). Toto, we're not in Kansas anymore: Measuring and using outcomes in behavioral health care. Clinical Psychology: Science & Practice, 5, 115-133.

Bryant, F. B. (2000). Assessing the validity of measurement. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding more multivariate statistics (pp. 99-146). Washington, DC: American Psychological Association.

Bryant, F. B., & Yarnold, P. R. (1997). Principle-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics (pp. 99-136). Washington, DC: American Psychological Association.

Burnlingame, G. M., & Lambert, M. J., Reisinger, C. W., Neff, W. M., & Mosier, J. (1995). Pragmatics of tracking mental health outcomes in a managed care setting. Journal of Mental Health Administration, 22, 226-236.

Byrne, B. M. (1995). One application of structural equation modeling from two perspectives: Exploring the OQS and LISREL strategies. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 138-157). Thousand Oaks, CA: Sage Publications.

Byrne, B. M. (2001). Structural equation modeling with AMOS: Basic concepts, applications, and programming. Mahwah, NJ: Lawrence Erlbaum.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 37-55). Thousand Oaks, CA: Sage Publications.

Cohen, J. & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L. J., (1990). Essentials of psychological testing (5th ed.). New York: Harper & Row.

Dornelas, E. A., Correll, R. E., Lothstein, L., Wilber, C., & Goethe, J. W. (1996). Designing and implementing outcome evaluations: Some guidelines for practitioners. Psychotherapy, 33, 237-245.

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. Psychological Assessment, 7, 248-260.

Goldfried, M. R., & Wolfe, B. E. (1998). Toward a more clinically valid approach to therapy research. Journal of Consulting and Clinical Psychology, 66, 143-150.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Gorsuch, R. L. (1988). Exploratory factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (2nd ed., pp. 231-258). New York: Plenum Press.

Groth-Marnat, G. (1990). Handbook of psychological assessment (2nd ed.). New York: John Wiley & Sons.

Hair, J. F., J. R., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). Multivariate data analysis (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Haynes, S. N., Nelson, K., & Blaine, D. D. (1999). Psychometric issues in assessment research. In Kendall, P. C., Butcher, J. N., & Holmbeck G. N. (Eds.). Handbook of Research Methods in Clinical Psychology (2nd ed.). New York: John Wiley & Sons.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. American Psychologist, 51, 1059-1064.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 158-176). Thousand Oaks, CA: Sage Publications.

Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 76-99). Thousand Oaks, CA: Sage Publications.

Johnson, L. D., & Shaha, S. (1996). Improving quality in psychotherapy. Psychotherapy, 33, 225-236.

Kazdin, A. E. (1998). Research design in clinical psychology (3rd ed.). Boston: Allyn and Bacon.

Kazdin, A. E. (1999). Overview of research design issues in clinical psychology. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.) Handbook of research methods in clinical psychology (2nd ed., pp. 3-30). New York: John Wiley & Sons.

Kline, R. B. (1998). Principles and practice of structural equation modeling. New York: Guilford Press.

Lambert, M. J. (Ed.). (2001). Patient-focused research [Special section]. Journal of Consulting and Clinical Psychology, 69 (2).

Lambert, M. J. (2001). Psychotherapy outcome and quality improvement: Introduction to the special section on patient-focused research. Journal of Consulting and Clinical Psychology, 69, 147-149.

Lambert, M. J., Burnlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the Outcome Questionnaire. Clinical Psychology and Psychotherapy, 3, 249-258.

Lambert, M. J., & Finch, A. E. (1999). The outcome questionnaire. In M. E. Maruish (Ed.), The use of psychological testing for treatment planning and outcome assessment (pp. 831-869. Mahwah, NJ: Lawrence Erlbaum Associates.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. Journal of Clinical and Consulting Psychology, 69, 159-172.

Lambert, M. J., Huefner, J. C., & Nace, D. K. (1997). The promise and problems of psychotherapy research in a managed care setting. Psychotherapy Research, 7, 321-332.

Lyons, J. S., Howard, K. I., O'Hahoney, M. T., & Lish, J. D. (1997). The measurement & management of clinical outcomes in mental health. New York: John Wiley & Sons.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. Psychological Bulletin, 111, 490-504.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519-530.

Messick, S. (1995) Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50, 741-749.

Mueller, R. M., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the outcome questionnaire: A confirmatory factor analysis. Journal of Personality Assessment, 70, 248-262.

Nebeker, R. S., Lambert, M. J., & Huefner, J. C. (1995). Ethnic differences on the Outcome Questionnaire. Psychological Reports, 77, 875-879.

Newsom, J. (2001, Winter). Practical approaches to dealing with nonnormal and categorical variables [On-line]. Available: www.ioa.pdx.edu/newsom/semclass/ho_estimate2.doc.

Nunally, J. C. & Bernstein, I. H. (1994). Psychometric Theory (3rd ed.). New York: McGraw-Hill.

OQ Systems, Inc. (1999). Therapeutic monitoring system (TMS): Implementation and operations manual.

Pallak, M. S., & Cummings, N. A. (1994). Outcomes research in managed behavioral health care: issues, strategies, and trends. In S. A. Shueman, W. G. Troy, and S. L. Mayhugh (Eds.), Managed behavioral health care: An industry perspective. Springfield, IL: Charles C Thomas Publisher.

Reckase, M. D. (1996). Test construction in the 1990s: Recent approaches every psychologist should know. Psychological Assessment, 8, 354-359.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. Psychological Assessment, 8, 350-353.

Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. Psychological Assessment, 7, 300-308.

Sonnanburg, K. (1996). Meaningful measurement in psychotherapy. Psychotherapy, 33, 160-170.

SPSS (1997a). SPSS Base 7.5 applications guide. Chicago: SPSS Inc.

SPSS (1997b). SPSS missing value analysis 7.5. Chicago: SPSS Inc.

Strube, M. J. (2000). Reliability and generalizability theory. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding more multivariate statistics (pp. 23-66). Washington, DC: American Psychological Association.

Thompson, B. (2000). Ten commandments of structural equation modeling. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding more multivariate statistics (pp. 261-283). Washington, DC: American Psychological Association.

Umphress, V. J., Lambert, M. J., Smart, D. W., Barlow, S. H., & Clouse, G. (1997). Concurrent and construct validity of the Outcome Questionnaire. Journal of Psychoeducational Assessment, 15, 40-55.

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome Questionnaire: Item sensitivity to change. Journal of Personality Assessment, 74, 242-261.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 56-75). Thousand Oaks, CA: Sage Publications.

Wells, M. G., Burlingame, G. M., Lambert, M. J., Hoag, M. J., & Hope, C. A. (1996). Conceptualization and measurement of patient change during psychotherapy: Development of the Outcome Questionnaire and Youth Outcome Questionnaire. Psychotherapy, 33, 275-283.

APPENDIX A: PROGRESS QUESTIONNAIRE

0=Never	1=Rarely	2=Sometimes	3=Frequently	4=Almost Always	**INSTRUCTIONS:** *Looking back over the last week, including today, help us understand how you have been feeling. Read each item carefully and* ***circle the number*** *best describes your current situation. For this questionnaire,* ***work is defined as employment, school, housework, volunteer work and so forth****.*
0	1	2	3	4	1. I get along well with others.
0	1	2	3	4	2. I tire quickly.
0	1	2	3	4	3. I feel no interest in things.
0	1	2	3	4	4. I feel stressed at work/school.
0	1	2	3	4	5. I blame myself for things.
0	1	2	3	4	6. I feel irritated.
0	1	2	3	4	7. I feel unhappy in my marriage/significant relationship.
0	1	2	3	4	8. I have thoughts of ending my life.
0	1	2	3	4	9. I feel weak.
0	1	2	3	4	10. I feel fearful.
0	1	2	3	4	11. After heavy drinking, I need a drink the next morning to get going (If you do not drink, mark “never”).
0	1	2	3	4	12. I find my work/school satisfying.
0	1	2	3	4	13. I am a happy person.
0	1	2	3	4	14. I work/study too much.
0	1	2	3	4	15. I feel worthless.
0	1	2	3	4	16. I am concerned about family troubles.
0	1	2	3	4	17. I have an unfulfilling sex life.
0	1	2	3	4	18. I feel lonely.
0	1	2	3	4	19. I have frequent arguments.
0	1	2	3	4	20. I feel loved and wanted.
0	1	2	3	4	21. I enjoy my spare time.
0	1	2	3	4	22. I have difficulty concentrating.
0	1	2	3	4	23. I feel hopeless about the future.
0	1	2	3	4	24. I like myself.
0	1	2	3	4	25. Disturbing thoughts come into my mind that I cannot get rid of.
0	1	2	3	4	26. I feel annoyed by people who criticize my drinking (or drug use) (if not applicable, mark “never”).
0	1	2	3	4	27. I have an upset stomach.
0	1	2	3	4	28. I am not working/studying as well as I used to.
0	1	2	3	4	29. My heart pounds too much.

0=Never	1=Rarely	2=Sometimes	3=Frequently	4=Almost Always	**INSTRUCTIONS:** *Looking back over the last week, including today, help us understand how you have been feeling. Read each item carefully and* ***circle the number*** *best describes your current situation. For this questionnaire,* ***work is defined as employment, school, housework, volunteer work and so forth****.*
0	1	2	3	4	30. I have trouble getting along with friends and close acquaintances.
0	1	2	3	4	31. I am satisfied with my life.
0	1	2	3	4	32. I have trouble at work/school because of my drinking or drug use (if not applicable, mark "never").
0	1	2	3	4	33. I feel that something bad is going to happen.
0	1	2	3	4	34. I have sore muscles.
0	1	2	3	4	35. I feel afraid of open spaces, of driving, or being on buses, subways, and so forth.
0	1	2	3	4	36. I feel nervous.
0	1	2	3	4	37. I feel my love relationships are full and complete.
0	1	2	3	4	38. I feel that I am not doing well at work/school.
0	1	2	3	4	39. I have too many disagreements at work/school.
0	1	2	3	4	40. I feel something is wrong with my mind.
0	1	2	3	4	41. I have trouble falling asleep or staying asleep.
0	1	2	3	4	42. I feel blue.
0	1	2	3	4	43. I am satisfied with my relationships with others.
0	1	2	3	4	44. I feel angry enough at work/school to do something I might regret.
0	1	2	3	4	45. I have headaches.
0	1	2	3	4	46. I have episodes where I do things to hurt myself physically.
0	1	2	3	4	47. I worry about my physical health problems.
0	1	2	3	4	48. I have crises where I call my therapist or psychiatrist.
0	1	2	3	4	49. I have thoughts of hurting other people.
0	1	2	3	4	50. I have days where I am less productive at work because of emotional distress.
0	1	2	3	4	51. I have a poor appetite or I find myself overeating.
0	1	2	3	4	52. I have been moving or speaking so slowly that other people have noticed.
0	1	2	3	4	53. I have been so fidgety or restless that other people have noticed.
0	1	2	3	4	54. I have had periods that lasted more than a few days where I felt speeded up, had lots of energy, and didn't need sleep.
0	1	2	3	4	55. I have excessive worry (that I can't control) occurring more days than not, about a number of events or activities.
0	1	2	3	4	56. I am taking medication for depression, anxiety, or other mental health problems.
0	1	2	3	4	57. I am taking my medication exactly as prescribed.
0	1	2	3	4	58. I see and hear things that other people do not see or hear and that really might not be there.

0=Never	1=Rarely	2=Sometimes	3=Frequently	4=Almost Always	**INSTRUCTIONS:** *Looking back over the last week, including today, help us understand how you have been feeling. Read each item carefully and* ***circle the number*** *best describes your current situation. For this questionnaire,* ***work is defined as employment, school, housework, volunteer work and so forth****.*
0	1	2	3	4	59. I have panic or anxiety attacks that come suddenly out of the blue.
0	1	2	3	4	60. I have upsetting thoughts, nightmares, or flashbacks about a stressful event that I experienced.

APPENDIX B: PQ CFA PARAMETER ESTIMATES

Table B1: PQ CFA Unstandardized Regression Weights

			Estimate	S.E.	C.R.	P
pq02	←	PQ-SD	1			
pq03	←	PQ-SD	1.376	0.176	7.816	.0001
pq05	←	PQ-SD	1.284	0.172	7.453	.0001
pq06	←	PQ-SD	1.218	0.158	7.697	.0001
pq08	←	PQ-SD	1.093	0.163	6.723	.0001
pq09	←	PQ-SD	1.715	0.206	8.342	.0001
pq10	←	PQ-SD	1.537	0.2	7.7	.0001
pq11t	←	PQ-SD	-0.07	0.026	-2.664	.008
pq13	←	PQ-SD	1.102	0.157	7.008	.0001
pq15	←	PQ-SD	1.768	0.21	8.434	.0001
pq22	←	PQ-SD	1.139	0.158	7.223	.0001
pq23	←	PQ-SD	1.707	0.209	8.18	.0001
pq24	←	PQ-SD	1.307	0.176	7.405	.0001
pq25	←	PQ-SD	1.571	0.2	7.851	.0001
pq27	←	PQ-SD	0.992	0.154	6.449	.0001
pq29	←	PQ-SD	1.256	0.18	6.963	.0001
pq31	←	PQ-SD	1.381	0.186	7.423	.0001
pq33	←	PQ-SD	1.614	0.201	8.032	.0001
pq34	←	PQ-SD	0.871	0.157	5.557	.0001

Table B1: PQ CFA Unstandardized Regression Weights (*continued*)

			Estimate	S.E.	C.R.	P
pq35	←	PQ-SD	0.866	0.163	5.318	.0001
pq36	←	PQ-SD	1.543	0.194	7.945	.0001
pq40	←	PQ-SD	1.743	0.217	8.024	.0001
pq41	←	PQ-SD	1.488	0.203	7.343	.0001
pq42	←	PQ-SD	1.647	0.196	8.402	.0001
pq45	←	PQ-SD	0.699	0.147	4.759	.0001
pq46	←	PQ-SD	0.606	0.114	5.303	.0001
pq47	←	PQ-SD	0.981	0.172	5.698	.0001
pq49	←	PQ-SD	0.536	0.12	4.472	.0001
pq51	←	PQ-SD	1.252	0.182	6.882	.0001
pq52	←	PQ-SD	1.001	0.16	6.255	.0001
pq53	←	PQ-SD	1.338	0.195	6.875	.0001
pq54	←	PQ-SD	0.777	0.146	5.314	.0001
pq55	←	PQ-SD	1.864	0.231	8.062	.0001
pq58	←	PQ-SD	0.398	0.111	3.573	.0001
pq59	←	PQ-SD	1.442	0.198	7.289	.0001
pq60	←	PQ-SD	1.542	0.21	7.351	.0001
pq01	←	PQ-IR	1			
pq07	←	PQ-IR	2.116	0.416	5.082	.0001
pq16	←	PQ-IR	1.11	0.278	3.989	.0001
pq17	←	PQ-IR	1.784	0.377	4.735	.0001

Table B1: PQ CFA Unstandardized Regression Weights (*continued*)

			Estimate	S.E.	C.R.	P
pq18	←	PQ-IR	2.6	0.436	5.959	.0001
pq19	←	PQ-IR	1.528	0.31	4.927	.0001
pq20	←	PQ-IR	2.633	0.438	6.012	.0001
pq26	←	PQ-IR	0.866	0.235	3.687	.0001
pq30	←	PQ-IR	1.942	0.338	5.745	.0001
pq37	←	PQ-IR	2.6	0.454	5.722	.0001
pq43	←	PQ-IR	2.252	0.383	5.873	.0001
pq04	←	PQ-SR	1			
pq12	←	PQ-SR	0.634	0.099	6.396	.0001
pq14	←	PQ-SR	0.307	0.092	3.346	.001
pq21	←	PQ-SR	0.701	0.099	7.101	.0001
pq28	←	PQ-SR	0.945	0.108	8.741	.0001
pq32t	←	PQ-SR	-0.053	0.018	-2.922	.003
pq38	←	PQ-SR	0.881	0.11	8.011	.0001
pq39	←	PQ-SR	0.666	0.09	7.368	.0001
pq44	←	PQ-SR	0.753	0.096	7.872	.0001
pq50	←	PQ-SR	0.986	0.103	9.542	.0001

Table B2: Standardized Regression Weights

			Estimate
pq02	←	PQ-SD	.494
pq03	←	PQ-SD	.67
pq05	←	PQ-SD	.613
pq06	←	PQ-SD	.651
pq08	←	PQ-SD	.516
pq09	←	PQ-SD	.768
pq10	←	PQ-SD	.651
pq11t	←	PQ-SD	- .168
pq13	←	PQ-SD	.552
pq15	←	PQ-SD	.787
pq22	←	PQ-SD	.58
pq23	←	PQ-SD	.735
pq24	←	PQ-SD	.606
pq25	←	PQ-SD	.676
pq27	←	PQ-SD	.484
pq29	←	PQ-SD	.546
pq31	←	PQ-SD	.609
pq33	←	PQ-SD	.708
pq34	←	PQ-SD	.393
pq35	←	PQ-SD	.371
pq36	←	PQ-SD	.692

Table B2: Standardized Regression Weights (*continued*)

			Estimate
pq40	←	PQ-SD	.706
pq41	←	PQ-SD	.597
pq42	←	PQ-SD	.78
pq45	←	PQ-SD	.323
pq46	←	PQ-SD	.37
pq47	←	PQ-SD	.407
pq49	←	PQ-SD	.3
pq51	←	PQ-SD	.536
pq52	←	PQ-SD	.463
pq53	←	PQ-SD	.535
pq54	←	PQ-SD	.371
pq55	←	PQ-SD	.713
pq58	←	PQ-SD	.232
pq59	←	PQ-SD	.59
pq60	←	PQ-SD	.598
pq01	←	PQ-IR	.383
pq07	←	PQ-IR	.471
pq16	←	PQ-IR	.311
pq17	←	PQ-IR	.411
pq18	←	PQ-IR	.72
pq19	←	PQ-IR	.443

Table B2: Standardized Regression Weights (*continued*)

			Estimate
pq20	←	PQ-IR	.744
pq26	←	PQ-IR	.278
pq30	←	PQ-IR	.638
pq37	←	PQ-IR	.631
pq43	←	PQ-IR	.685
pq04	←	PQ-SR	.653
pq12	←	PQ-SR	.438
pq14	←	PQ-SR	.222
pq21	←	PQ-SR	.492
pq28	←	PQ-SR	.627
pq32t	←	PQ-SR	- .193
pq38	←	PQ-SR	.565
pq39	←	PQ-SR	.513
pq44	←	PQ-SR	.554
pq50	←	PQ-SR	.7

Table B3: PQ Subscale Covariances

			Estimate	S.E.	C.R.	P
PQ-SD	↔	PQ-IR	.132	0.028	4.649	.0001
PQ-SD	↔	PQ-SR	.318	0.055	5.783	.0001

Table B3: PQ Subscale Covariances (*continued*)

			Estimate	S.E.	C.R.	P
PQ-IR	↔	PQ-SR	.191	0.04	4.805	.0001

Table B4: PQ Subscale Correlations

			Estimate
PQ-SD	↔	PQ-IR	.761
PQ-SD	↔	PQ-SR	.694
PQ-IR	↔	PQ-SR	.684

Table B5: PQ CFA Variances

	Estimate	S.E.	C.R.	P
PQ-SD	0.285	0.066	4.299	.0001
PQ-IR	0.105	0.034	3.106	.002
PQ-SR	0.74	0.131	5.651	.0001
e02	0.884	0.076	11.598	.0001
e03	0.661	0.058	11.336	.0001
e05	0.778	0.068	11.449	.0001
e06	0.576	0.051	11.379	.0001
e08	0.937	0.081	11.576	.0001
e09	0.583	0.053	11.007	.0001

Table B5: PQ CFA Variances (*continued*)

	Estimate	S.E.	C.R.	P
e10	0.913	0.08	11.378	.0001
e11	0.048	0.004	11.753	.0001
e13	0.791	0.069	11.537	.0001
e15	0.545	0.05	10.904	.0001
e22	0.727	0.063	11.499	.0001
e23	0.704	0.063	11.144	.0001
e24	0.837	0.073	11.461	.0001
e25	0.835	0.074	11.322	.0001
e27	0.913	0.079	11.606	.0001
e29	1.059	0.092	11.544	.0001
e31	0.921	0.08	11.456	.0001
e33	0.739	0.066	11.236	.0001
e34	1.181	0.101	11.672	.0001
e35	1.337	0.114	11.684	.0001
e36	0.736	0.065	11.28	.0001
e40	0.869	0.077	11.241	.0001
e41	1.137	0.099	11.475	.0001
e42	0.496	0.045	10.943	.0001
e45	1.194	0.102	11.707	.0001
e46	0.661	0.057	11.685	.0001

Table B5: PQ CFA Variances (*continued*)

	Estimate	S.E.	C.R.	P
e47	1.383	0.119	11.664	.0001
e49	0.828	0.071	11.716	.0001
e51	1.109	0.096	11.555	.0001
e52	1.044	0.09	11.624	.0001
e53	1.271	0.11	11.556	.0001
e54	1.08	0.092	11.684	.0001
e55	0.954	0.085	11.219	.0001
e58	0.794	0.068	11.739	.0001
e59	1.111	0.097	11.486	.0001
e60	1.214	0.106	11.473	.0001
e01	0.612	0.053	11.484	.0001
e07	1.649	0.146	11.295	.0001
e16	1.214	0.105	11.592	.0001
e17	1.65	0.144	11.432	.0001
e18	0.661	0.066	9.966	.0001
e19	1.007	0.089	11.364	.0001
e20	0.587	0.061	9.684	.0001
e26	0.942	0.081	11.63	.0001
e30	0.577	0.054	10.621	.0001
e37	1.075	0.101	10.665	.0001
e43	0.604	0.059	10.292	.0001

Table B5: PQ CFA Variances (*continued*)

	Estimate	S.E.	C.R.	P
e04	0.996	0.099	10.105	.0001
e12	1.253	0.111	11.24	.0001
e14	1.349	0.116	11.654	.0001
e21	1.136	0.103	11.057	.0001
e28	1.021	0.099	10.32	.0001
e32	0.053	0.005	11.683	.0001
e38	1.225	0.114	10.723	.0001
e39	0.918	0.084	10.973	.0001
e44	0.949	0.088	10.783	.0001
e50	0.748	0.078	9.608	.0001

APPENDIX C: PQ PCA FACTOR SCORES

Table C1: PQ Pattern Matrix

	1	2	3	4	5	6	7	8	9	10
PQ25	.792	-.143	-.084	.075	-.022	-.072	-.13	.246	.002	.115
PQ23	.749	.204	-.091	.071	-.106	-.053	.074	.038	-.126	.015
PQ08	.721	.011	-.173	-.143	-.084	.2	.019	.259	-.015	-.092
PQ05	.702	-.055	-.035	-.103	.086	-.01	.203	-.184	.129	-.034
PQ40	.689	.016	-.194	.034	.171	.004	-.084	.16	.086	.136
PQ42	.684	.02	.099	.024	.127	.044	-.034	.039	-.108	-.007
PQ15	.682	.15	-.127	.135	.049	.058	.093	-.05	-.005	.039
PQ33	.645	-.048	-.081	.33	-.058	.002	-.023	.087	.054	-.103
PQ06	.613	.114	.19	-.201	-.075	-.143	-.021	.064	.28	.086
PQ21	.584	.328	.006	-.175	-.003	-.103	-.09	.034	-.032	.123
PQ09	.538	.056	.21	.147	-.014	.007	.107	-.099	-.125	.149
PQ55	.537	-.1	.115	.306	.018	-.051	.016	.134	-.049	-.066
PQ24	.516	.422	-.086	-.018	-.039	.052	-.005	.003	-.064	-.02
PQ22	.516	.05	.031	-.013	.168	-.109	-.184	-.084	.039	.292
PQ03	.503	.196	.065	.006	.098	-.017	-.015	-.209	.002	.259
PQ18	.406	.207	-.045	.031	.203	.046	.298	-.097	.131	-.071
PQ60	.353	.012	.219	.282	-.06	.096	-.115	-.067	.219	-.197
PQ01	-.157	.741	-.021	.169	-.191	-.039	-.256	-.059	.287	.117
PQ13	.239	.667	.133	.049	-.074	-.01	-.205	.069	-.181	-.033

Table C1: PQ Pattern Matrix (*continued*)

	1	2	3	4	5	6	7	8	9	10
PQ43	.154	.624	-.129	.03	-.077	.086	.135	.065	.114	.126
PQ37	.108	.607	-.044	.013	-.011	-.145	.299	.126	-.036	-.102
PQ20	.123	.552	-.025	.021	.13	.084	.222	.009	.05	-.083
PQ30	-.127	.546	-.044	.236	.045	.137	.064	-.036	.377	.165
PQ31	.438	.536	-.035	-.052	-.081	-.003	.16	.101	-.132	-.053
PQ12	-.132	.515	.29	-.168	.309	-.03	-.075	.094	-.183	-.221
PQ45	-.079	.022	.833	-.272	-.003	.029	-.065	.03	.181	-.017
PQ27	-.075	-.035	.75	-.026	-.062	.092	.143	.2	.025	.182
PQ34	-.222	.032	.679	.219	-.23	.023	.194	.111	.038	.154
PQ02	.11	.085	.537	.02	-.009	.061	.001	-.202	-.106	.261
PQ29	.207	-.106	.425	.299	-.038	.059	-.038	.046	.02	-.127
PQ51	.39	-.124	.418	-.128	.123	.055	-.054	-.186	.079	.13
PQ35	-.093	.105	-.244	.821	-.031	-.055	.15	-.078	.083	.137
PQ59	.211	.072	.055	.632	.038	-.008	-.259	-.093	.024	-.004
PQ36	.188	.128	.161	.565	.084	-.04	.051	-.075	-.131	.01
PQ10	.418	-.052	-.09	.541	.075	-.035	.16	-.129	-.123	.007
PQ47	-.113	-.108	.388	.403	-.02	-.039	.097	.154	-.175	.342
PQ38	.141	-.135	-.33	.049	.915	.104	-.072	-.107	.	.047
PQ28	.006	-.053	.034	-.015	.681	.049	.099	.174	-.127	.379
PQ04	.057	.179	.156	.03	.584	-.068	-.008	-.044	-.015	-.299
PQ50	.231	-.124	.12	-.074	.581	.018	.059	.148	.022	.074

Table C1: PQ Pattern Matrix (*continued*)

	1	2	3	4	5	6	7	8	9	10
PQ39	-.305	.155	.033	.055	.453	-.075	.026	.335	.376	-.04
PQ11	-.122	-.02	.144	.09	.092	.852	-.07	-.176	-.009	-.096
PQ32	.101	-.053	.023	-.146	.032	.837	.066	-.052	-.04	.066
PQ26	-.13	.081	.072	-.078	.054	.755	.078	.073	.059	.023
PQ17	-.172	.18	-.018	.147	.068	.091	.741	-.019	.02	.05
PQ07	.236	-.023	.186	-.143	-.087	.006	.601	-.035	.299	-.141
PQ16	.246	-.212	.153	.133	-.091	-.158	.514	-.118	.272	-.037
PQ49	.105	.098	.144	-.182	-.04	-.083	-.043	.761	-.002	.21
PQ44	.025	.097	.008	.059	.246	-.064	-.031	.629	.16	-.018
PQ46	.305	.075	-.003	-.01	-.242	.367	-.127	.411	.046	.039
PQ19	.046	.08	.148	-.114	-.121	-.013	.274	-.014	.78	.05
PQ54	.062	-.172	-.024	.152	.172	.047	.019	.233	.538	.111
PQ52	.128	.007	.146	.108	-.047	.013	.005	.169	.037	.628
PQ53	.235	-.024	.179	.026	.09	-.022	-.135	.042	.251	.542

Table C2: PQ Structure Matrix

	1	2	3	4	5	6	7	8	9	10
PQ15	.799	.466	.347	.453	.354	.234	.31	.131	.16	.187
PQ42	.786	.366	.495	.428	.432	.194	.205	.178	.118	.123
PQ23	.771	.467	.312	.382	.222	.135	.318	.146	.035	.157

Table C2: PQ Structure Matrix (*continued*)

	1	2	3	4	5	6	7	8	9	10
PQ09	.745	.345	.549	.481	.317	.124	.333	.	.011	.3
PQ25	.724	.207	.317	.433	.253	.167	.06	.312	.192	.169
PQ40	.724	.373	.273	.38	.412	.235	.1	.293	.28	.204
PQ33	.722	.251	.36	.601	.249	.208	.136	.267	.241	-.007
PQ55	.701	.211	.499	.613	.332	.132	.19	.263	.142	.022
PQ05	.692	.265	.335	.223	.326	.068	.352	-.068	.192	.091
PQ06	.682	.414	.467	.209	.289	.046	.108	.134	.413	.14
PQ03	.662	.44	.414	.298	.359	.086	.194	-.111	.119	.396
PQ18	.658	.496	.358	.315	.467	.166	.436	.098	.234	.051
PQ24	.621	.601	.241	.245	.246	.211	.202	.16	.119	.11
PQ21	.614	.531	.295	.142	.275	.059	.123	.093	.131	.219
PQ08	.605	.305	.117	.197	.142	.387	.171	.393	.166	-.018
PQ22	.582	.301	.367	.271	.371	.018	.	-.037	.174	.369
PQ60	.576	.242	.51	.541	.26	.22	-.032	.164	.415	-.112
PQ43	.446	.727	.131	.193	.192	.281	.25	.233	.225	.212
PQ13	.494	.721	.346	.269	.245	.152	.019	.212	.077	.082
PQ20	.483	.698	.264	.224	.389	.222	.352	.213	.191	.023
PQ31	.612	.696	.255	.216	.241	.174	.374	.23	.032	.077
PQ37	.393	.674	.179	.15	.25	.013	.426	.223	.049	-.026
PQ30	.363	.652	.233	.343	.297	.322	.086	.209	.483	.214
PQ01	.148	.641	.094	.157	.03	.119	-.231	.098	.404	.142

Table C2: PQ Structure Matrix (*continued*)

	1	2	3	4	5	6	7	8	9	10
PQ12	.19	.539	.345	.014	.457	.002	.058	.19	.028	-.171
PQ27	.379	.193	.718	.334	.265	.181	.206	.225	.138	.207
PQ45	.255	.176	.708	.084	.281	.034	-.045	.062	.309	-.009
PQ34	.26	.143	.622	.417	.085	.099	.214	.135	.086	.181
PQ29	.5	.118	.614	.566	.266	.155	.058	.19	.201	-.058
PQ02	.424	.238	.606	.283	.245	.062	.15	-.167	-.016	.374
PQ51	.542	.153	.591	.236	.363	.074	.078	-.116	.195	.223
PQ59	.507	.221	.435	.735	.281	.137	-.128	.112	.23	.084
PQ36	.596	.32	.541	.727	.375	.089	.23	.095	.036	.136
PQ35	.265	.155	.11	.688	.098	.091	.176	.077	.089	.19
PQ10	.628	.182	.36	.673	.296	.08	.334	.021	-.025	.143
PQ47	.282	.035	.482	.532	.175	.064	.187	.134	-.117	.37
PQ38	.291	.133	.1	.177	.784	.116	.032	.053	.137	.083
PQ04	.383	.369	.449	.249	.711	-.036	.115	.128	.19	-.237
PQ28	.358	.249	.328	.236	.708	.147	.238	.209	-.007	.396
PQ50	.492	.23	.443	.268	.702	.121	.185	.243	.189	.101
PQ39	.104	.312	.215	.18	.536	.093	-.038	.461	.496	-.137
PQ32	.215	.132	.044	.06	.06	.808	.103	.187	.075	.17
PQ11	.133	.095	.155	.215	.116	.778	-.063	.14	.153	.013
PQ26	.128	.223	.067	.09	.106	.771	.067	.322	.186	.081

Table C2: PQ Structure Matrix (*continued*)

	1	2	3	4	5	6	7	8	9	10
PQ46	.358	.265	.102	.247	-.026	.576	-.064	.558	.244	.053
PQ17	.22	.297	.125	.186	.189	.131	.746	.062	-.088	.122
PQ07	.419	.213	.318	.104	.167	.06	.589	.046	.231	-.083
PQ16	.403	.007	.344	.289	.129	-.108	.502	-.086	.167	.011
PQ44	.305	.334	.234	.305	.429	.213	-.002	.719	.366	-.103
PQ49	.241	.289	.183	.108	.164	.198	.015	.7	.159	.112
PQ19	.303	.282	.274	.095	.144	.128	.142	.119	.729	.016
PQ54	.305	.103	.232	.343	.322	.245	-.057	.377	.611	.042
PQ52	.377	.213	.306	.32	.149	.192	.102	.141	.08	.641
PQ53	.47	.242	.414	.309	.31	.148	-.05	.072	.336	.553

APPENDIX D: OQ-45.2 CFA PARAMETER ESTIMATES

Table D1: OQ-45.2 CFA Unstandardized Regression Weights

			Estimate	S.E.	C.R.	P
OQ02	←	SD	1			
OQ03	←	SD	1.321	0.126	10.471	.0001
OQ05	←	SD	1.293	0.122	10.564	.0001
OQ06	←	SD	1.205	0.113	10.674	.0001
OQ08	←	SD	1.05	0.104	10.061	.0001
OQ09	←	SD	1.548	0.141	11.006	.0001
OQ10	←	SD	1.534	0.143	10.76	.0001
OQ11t	←	SD	-0.039	0.016	-2.465	.014
OQ13	←	SD	1.197	0.115	10.379	.0001
OQ15	←	SD	1.815	0.154	11.755	.0001
OQ22	←	SD	1.174	0.115	10.166	.0001
OQ23	←	SD	1.759	0.153	11.479	.0001
OQ24	←	SD	1.445	0.131	11.073	.0001
OQ25	←	SD	1.379	0.135	10.202	.0001
OQ27	←	SD	1.108	0.123	9.025	.0001
OQ29	←	SD	1.346	0.133	10.147	.0001
OQ31	←	SD	1.339	0.128	10.442	.0001

Table D1: OQ-45.2 CFA Unstandardized Regression Weights (*continued*)

			Estimate	S.E.	C.R.	P
OQ33	←	SD	1.453	0.138	10.542	.0001
OQ34	←	SD	0.749	0.112	6.672	.0001
OQ35	←	SD	0.752	0.095	7.908	.0001
OQ36	←	SD	1.318	0.128	10.29	.0001
OQ40	←	SD	1.781	0.16	11.12	.0001
OQ41	←	SD	1.338	0.143	9.356	.0001
OQ42	←	SD	1.749	0.15	11.672	.0001
OQ45	←	SD	1.08	0.127	8.504	.0001
OQ01	←	IR	1			
OQ07	←	IR	3.464	0.5	6.925	.0001
OQ16	←	IR	1.837	0.316	5.822	.0001
OQ17	←	IR	3	0.46	6.518	.0001
OQ18	←	IR	3.926	0.537	7.314	.0001
OQ19	←	IR	2.295	0.358	6.409	.0001
OQ20	←	IR	3.468	0.472	7.342	.0001
OQ26t	←	IR	-0.098	0.045	-2.18	.029
OQ30	←	IR	1.507	0.259	5.827	.0001
OQ37	←	IR	3.447	0.49	7.038	.0001
OQ43	←	IR	2.841	0.397	7.148	.0001

Table D1: OQ-45.2 CFA Unstandardized Regression Weights (*continued*)

			Estimate	S.E.	C.R.	P
OQ04	←	SR	1			
OQ12	←	SR	0.986	0.095	10.339	.0001
OQ14	←	SR	0.048	0.08	0.592	.554
OQ21	←	SR	0.973	0.095	10.247	.0001
OQ28	←	SR	1.23	0.107	11.472	.0001
OQ32t	←	SR	-0.037	0.013	-2.904	.004
OQ38	←	SR	1.213	0.103	11.823	.0001
OQ39	←	SR	0.745	0.08	9.314	.0001
OQ44	←	SR	0.718	0.078	9.157	.0001

Table D2: Standardized Regression Weights

			Estimate
OQ02	←	SD	.528
OQ03	←	SD	.657
OQ05	←	SD	.667
OQ06	←	SD	.679
OQ08	←	SD	.615
OQ09	←	SD	.716
OQ10	←	SD	.688

Table D2: Standardized Regression Weights (*continued*)

			Estimate
OQ11t	←	SD	-.12
OQ13	←	SD	.647
OQ15	←	SD	.814
OQ22	←	SD	.625
OQ23	←	SD	.776
OQ24	←	SD	.724
OQ25	←	SD	.629
OQ27	←	SD	.522
OQ29	←	SD	.623
OQ31	←	SD	.654
OQ33	←	SD	.664
OQ34	←	SD	.352
OQ35	←	SD	.436
OQ36	←	SD	.638
OQ40	←	SD	.73
OQ41	←	SD	.55
OQ42	←	SD	.802
OQ45	←	SD	.48
OQ01	←	IR	.361

Table D2: Standardized Regression Weights (*continued*)

			Estimate
OQ07	←	IR	.623
OQ16	←	IR	.407
OQ17	←	IR	.523
OQ18	←	IR	.765
OQ19	←	IR	.502
OQ20	←	IR	.778
OQ26t	←	IR	-.113
OQ30	←	IR	.408
OQ37	←	IR	.658
OQ43	←	IR	.696
OQ04	←	SR	.618
OQ12	←	SR	.585
OQ14	←	SR	.03
OQ21	←	SR	.579
OQ28	←	SR	.67
OQ32t	←	SR	-.148
OQ38	←	SR	.698
OQ39	←	SR	.515
OQ44	←	SR	.505

Table D3: OQ-45.2 Subscale Covariances

			Estimate	S.E.	C.R.	P
SD	↔	SR	0.294	0.037	7.966	.0001
SD	↔	IR	0.101	0.017	5.876	.0001
IR	↔	SR	0.099	0.017	5.733	.0001

Table D4: OQ-45.2 Correlations

			Estimate
SD	↔	SR	.857
SD	↔	IR	.785
IR	↔	SR	.627

Table D5: OQ-45.2 CFA Variances

	Estimate	S.E.	C.R.	P
SD	0.279	0.047	5.913	.0001
IR	0.06	0.016	3.736	.0001
SR	0.422	0.062	6.82	.0001
e02	0.722	0.049	14.721	.0001
e03	0.643	0.044	14.469	.0001
e05	0.583	0.04	14.44	.0001
e06	0.474	0.033	14.404	.0001

Table D5: OQ-45.2 CFA Variances (*continued*)

	Estimate	S.E.	C.R.	P
e08	0.506	0.035	14.571	.0001
e09	0.633	0.044	14.267	.0001
e10	0.729	0.051	14.372	.0001
e11	0.029	0.002	14.973	.0001
e13	0.555	0.038	14.495	.0001
e15	0.468	0.034	13.649	.0001
e22	0.598	0.041	14.548	.0001
e23	0.571	0.041	13.956	.0001
e24	0.527	0.037	14.233	.0001
e25	0.811	0.056	14.539	.0001
e27	0.916	0.062	14.73	.0001
e29	0.796	0.055	14.552	.0001
e31	0.671	0.046	14.477	.0001
e33	0.746	0.052	14.448	.0001
e34	1.102	0.074	14.887	.0001
e35	0.674	0.045	14.824	.0001
e36	0.707	0.049	14.518	.0001
e40	0.774	0.055	14.207	.0001
e41	1.153	0.078	14.689	.0001
e42	0.473	0.034	13.757	.0001
e45	1.087	0.074	14.78	.0001

Table D5: OQ-45.2 CFA Variances (*continued*)

	Estimate	S.E.	C.R.	P
e01	0.397	0.027	14.701	.0001
e07	1.128	0.082	13.785	.0001
e16	1.011	0.069	14.609	.0001
e17	1.421	0.1	14.272	.0001
e18	0.65	0.053	12.294	.0001
e19	0.932	0.065	14.349	.0001
e20	0.466	0.039	12.055	.0001
e26	0.044	0.003	14.959	.0001
e30	0.677	0.046	14.608	.0001
e37	0.929	0.069	13.54	.0001
e43	0.511	0.039	13.199	.0001
e04	0.681	0.05	13.516	.0001
e12	0.787	0.057	13.752	.0001
e14	1.056	0.071	14.981	.0001
e21	0.793	0.057	13.794	.0001
e28	0.785	0.06	13.046	.0001
e32	0.026	0.002	14.931	.0001
e38	0.655	0.052	12.714	.0001
e39	0.649	0.046	14.133	.0001
e44	0.636	0.045	14.179	.0001

APPENDIX E: OQ-45.2 PCA FACTOR SCORES

Table E1: OQ-45.2 Pattern Matrix

	1	2	3	4	5	6	7	8	9
OQ24	.858	.031	-.15	-.051	-.054	.005	.089	.225	-.065
OQ13	.834	-.044	-.022	.02	-.11	-.002	.022	.322	.048
OQ31	.801	-.114	.036	-.066	.145	.012	.005	.048	.126
OQ21	.709	.053	-.039	.011	.075	-.005	-.123	.143	-.016
OQ03	.661	-.051	.015	.26	-.173	.009	.083	-.009	-.023
OQ43	.641	.01	.054	-.021	.151	-.041	-.045	.358	.057
OQ15	.541	.365	-.025	-.035	-.017	-.032	.201	-.019	-.066
OQ23	.524	.353	-.01	-.128	.039	-.039	.166	-.008	-.035
OQ20	.513	-.07	-.088	.03	.506	.026	.016	.165	.083
OQ42	.484	.21	.04	.082	.066	-.024	.112	-.208	-.013
OQ06	.461	-.13	.192	.183	-.094	-.067	.097	-.107	.423
OQ10	.067	.833	.082	-.165	-.029	-.11	-.059	-.06	.002
OQ36	.027	.737	-.05	.082	-.02	-.012	-.068	.027	.002
OQ33	.114	.649	.137	.	-.029	.014	-.094	.061	.02
OQ35	-.138	.621	-.299	.257	-.013	-.099	.192	.173	-.055
OQ29	-.11	.575	.238	.161	-.028	.06	-.107	.101	.059
OQ25	.017	.566	.028	-.143	-.021	.075	.317	.	.086
OQ05	.3	.456	.112	-.137	-.031	-.038	.007	-.202	.094
OQ38	.051	.105	.789	-.037	.033	.028	.005	.032	-.209

Table E1: OQ-45.2 Pattern Matrix (*continued*)

	1	2	3	4	5	6	7	8	9
OQ04	.053	-.07	.73	.052	-.081	-.05	.039	-.001	.109
OQ39	-.249	.018	.677	-.071	.027	.014	.238	.353	.141
OQ28	-.103	.224	.647	.195	.197	.014	-.098	-.046	-.182
OQ22	.23	.23	.264	.247	-.047	.063	-.029	-.134	-.175
OQ02	.315	-.223	.034	.773	-.042	.036	-.24	-.004	.006
OQ34	-.255	.194	-.045	.671	.133	-.026	.066	.183	-.166
OQ45	-.039	-.076	.056	.671	-.085	-.091	.057	-.039	.31
OQ27	-.137	.285	.026	.533	.042	-.02	.011	.04	.248
OQ09	.201	.235	-.002	.466	.041	.117	.029	-.045	-.054
OQ17	-.118	-.088	.066	.037	.789	.005	.018	-.102	.139
OQ07	-.067	.068	-.085	.025	.638	-.006	-.032	-.126	.483
OQ37	.303	-.061	.043	-.05	.633	.022	-.04	.082	.013
OQ18	.325	.292	.06	-.105	.405	-.046	-.038	-.079	.003
OQ26	-.105	-.069	-.068	.016	.075	.849	.173	-.068	.051
OQ32	.043	-.241	.078	.053	.061	.82	.094	-.04	-.058
OQ11	.013	.306	-.027	-.171	-.195	.642	-.26	.15	.139
OQ44	-.018	-.026	.281	-.143	-.035	-.019	.742	.071	.093
OQ08	.259	-.029	-.142	.036	.036	.106	.715	-.174	-.066
OQ40	.277	.275	-.006	.046	-.093	.036	.42	-.066	-.016
OQ01	.455	-.102	.049	-.005	-.009	-.039	-.103	.771	-.103
OQ30	.203	.24	.051	.154	-.127	.051	.001	.585	.151

Table E1: OQ-45.2 Pattern Matrix (*continued*)

	1	2	3	4	5	6	7	8	9
OQ19	.004	-.065	-.001	.058	.276	.047	.102	.103	.721
OQ16	.072	.41	-.234	.039	.054	.047	-.216	-.148	.478

Table E2: OQ-45.2 Structure Matrix

	1	2	3	4	5	6	7	8	9
OQ31	.804	.364	.338	.209	.419	.106	.212	-.051	.262
OQ15	.755	.69	.401	.395	.291	.115	.476	-.094	.093
OQ24	.753	.403	.237	.221	.265	.111	.312	.113	.073
OQ13	.735	.371	.346	.271	.193	.104	.288	.207	.184
OQ42	.728	.601	.404	.442	.322	.087	.348	-.285	.125
OQ23	.716	.633	.359	.289	.328	.103	.414	-.077	.119
OQ03	.695	.427	.405	.501	.09	.084	.326	-.11	.096
OQ21	.693	.385	.259	.243	.344	.094	.122	.016	.113
OQ43	.66	.368	.336	.213	.389	.084	.221	.258	.185
OQ06	.619	.367	.491	.417	.092	-.005	.292	-.166	.517
OQ18	.617	.534	.284	.214	.595	.084	.176	-.148	.13
OQ10	.46	.773	.352	.285	.2	.05	.248	-.129	.149
OQ36	.404	.735	.281	.41	.189	.134	.242	-.05	.118
OQ33	.483	.725	.431	.388	.196	.167	.253	-.011	.16
OQ25	.392	.665	.34	.257	.165	.213	.523	.023	.194

Table E2: OQ-45.2 Structure Matrix (*continued*)

	1	2	3	4	5	6	7	8	9
OQ29	.326	.659	.48	.468	.118	.198	.24	.057	.168
OQ05	.576	.615	.377	.269	.205	.076	.243	-.264	.232
OQ35	.172	.58	.064	.416	.106	.022	.385	.133	-.006
OQ38	.411	.426	.818	.369	.125	.145	.323	.049	-.069
OQ04	.345	.292	.764	.356	-.024	.027	.288	.017	.21
OQ28	.392	.527	.734	.53	.266	.139	.243	-.059	-.06
OQ39	.082	.233	.663	.181	.	.117	.444	.452	.204
OQ02	.419	.247	.328	.723	.068	.058	.008	-.137	.049
OQ09	.537	.615	.4	.679	.232	.223	.326	-.121	.043
OQ45	.232	.293	.341	.662	-.052	-.059	.23	-.088	.321
OQ34	.093	.385	.225	.657	.136	.055	.281	.147	-.154
OQ27	.285	.535	.357	.645	.124	.075	.271	-.011	.295
OQ22	.53	.557	.53	.558	.138	.162	.27	-.197	-.051
OQ17	.219	.148	.07	.063	.737	.064	.032	-.083	.151
OQ37	.489	.255	.163	.087	.728	.118	.086	.041	.09
OQ20	.638	.338	.177	.186	.683	.136	.185	.091	.177
OQ07	.277	.256	.02	.069	.647	.056	.003	-.137	.498
OQ26	.033	.127	.032	.038	.119	.836	.173	.041	.04
OQ32	.109	.054	.139	.075	.114	.8	.126	.048	-.053
OQ11	.034	.202	.029	-.086	-.067	.66	-.122	.157	.174
OQ44	.245	.302	.467	.174	.015	.069	.79	.203	.151

Table E2: OQ-45.2 Structure Matrix (*continued*)

	1	2	3	4	5	6	7	8	9
OQ08	.446	.387	.219	.298	.182	.172	.726	-.106	-.007
OQ40	.539	.596	.386	.41	.124	.15	.609	-.066	.095
OQ01	.251	.039	.171	.025	.115	.05	.108	.697	-.037
OQ30	.315	.406	.344	.314	.04	.177	.306	.53	.235
OQ19	.245	.201	.164	.101	.309	.1	.166	.118	.731
OQ16	.278	.397	-.016	.143	.193	.098	-.093	-.228	.52

APPENDIX F: OQ-45.2 VALIDATION OF NINE FACTOR STRUCTURE PARAMETER ESTIMATES

Table F1: OQ-45.2 Validation Unstandardized Regression Weights

			Estimate	S.E.	C.R.
OQ24	←	Depression	1		
OQ03	←	Depression	0.946	0.071	13.319
OQ43	←	Depression	0.856	0.068	12.549
OQ15	←	Depression	1.201	0.078	15.313
OQ23	←	Depression	1.225	0.081	15.113
OQ20	←	Depression	0.885	0.071	12.546
OQ42	←	Depression	1.175	0.077	15.292
OQ13	←	Depression	0.796	0.064	12.518
OQ31	←	Depression	1.101	0.074	14.97
OQ21	←	Depression	0.872	0.077	11.361
OQ36	←	Anxiety	1		
OQ33	←	Anxiety	0.997	0.079	12.586
OQ29	←	Anxiety	0.918	0.08	11.519
OQ05	←	Anxiety	0.769	0.072	10.696
OQ10	←	Anxiety	1.116	0.083	13.393
OQ35	←	Anxiety	0.392	0.05	7.877
OQ25	←	Anxiety	1.067	0.088	12.125

Table F1: OQ-45.2 Validation Unstandardized Regression Weights (*continued*)

			Estimate	S.E.	C.R.
OQ38	←	Role	1		
OQ39	←	Role	0.463	0.052	8.93
OQ04	←	Role	0.741	0.06	12.378
OQ28	←	Role	1.112	0.069	16.146
OQ45	←	Somatic	1		
OQ34	←	Somatic	0.883	0.104	8.489
OQ02	←	Somatic	0.912	0.098	9.338
OQ09	←	Somatic	1.308	0.117	11.215
OQ07	←	LoveRelat	1		
OQ17	←	LoveRelat	1.065	0.118	9.038
OQ37	←	LoveRelat	1.029	0.104	9.941
OQ18	←	LoveRelat	1.208	0.114	10.623
OQ32t	←	Alcohol	1		
oQ26t	←	Alcohol	1.394	0.092	15.133
oQ11t	←	Alcohol	0.754	0.062	12.235
OQ08	←	Danger	1		
OQ40	←	Danger	2.149	0.196	10.975
OQ44	←	Danger	1.064	0.128	8.283
OQ30	←	Relationships	1		

Table F1: OQ-45.2 Validation Unstandardized Regression Weights (*continued*)

			Estimate	S.E.	C.R.
OQ01	←	Relationships	0.573	0.088	6.511
OQ19	←	Conflict	1		
OQ16	←	Conflict	0.755	0.11	6.845
OQ27	←	Somatic	1.038	0.11	9.43
OQ22	←	Somatic	1.059	0.102	10.391
OQ06	←	Conflict	1.02	0.101	10.051

Table F2: Standardized Regression Weights

			Estimate
OQ24	←	Depression	.698
OQ03	←	Depression	.663
OQ43	←	Depression	.623
OQ15	←	Depression	.767
OQ23	←	Depression	.756
OQ20	←	Depression	.623
OQ42	←	Depression	.766
OQ13	←	Depression	.622
OQ31	←	Depression	.749
OQ21	←	Depression	.563
OQ36	←	Anxiety	.672

Table F2: Standardized Regression Weights *(continued)*

			Estimate
OQ33	←	Anxiety	.675
OQ29	←	Anxiety	.611
OQ05	←	Anxiety	.563
OQ10	←	Anxiety	.726
OQ35	←	Anxiety	.406
OQ25	←	Anxiety	.647
OQ38	←	Role	.779
OQ39	←	Role	.448
OQ04	←	Role	.611
OQ28	←	Role	.797
OQ45	←	Somatic	.549
OQ34	←	Somatic	.497
OQ02	←	Somatic	.568
OQ09	←	Somatic	.767
OQ07	←	LoveRelat	.559
OQ17	←	LoveRelat	.565
OQ37	←	LoveRelat	.655
OQ18	←	LoveRelat	.74
OQ32t	←	Alcohol	.815
OQ26t	←	Alcohol	.933
OQ11t	←	Alcohol	.567

Table F2: Standardized Regression Weights (*continued*)

			Estimate
OQ08	←	Danger	.536
OQ40	←	Danger	.804
OQ44	←	Danger	.496
OQ30	←	Relationships	.684
OQ01	←	Relationships	.467
OQ19	←	Conflict	.594
OQ16	←	Conflict	.402
OQ27	←	Somatic	.576
OQ22	←	Somatic	.67
OQ06	←	Conflict	.695

Table F3: OQ-45.2 Validation Subscale Covariances

			Estimate	S.E.	C.R.	P
Relationships	↔	Conflict	0.18	0.03	5.279	.0001
Danger	↔	Relationships	0.169	0.03	6.306	.0001
Alcohol	↔	Danger	-0.003	0	-0.91	.361
LoveRelat	↔	Alcohol	-0.008	0.01	-1.39	.164
Somatic	↔	LoveRelat	0.218	0.04	5.606	.0001
Role	↔	Somatic	0.435	0.05	8.279	.0001
Anxiety	↔	Role	0.43	0.05	8.615	.0001

Table F3: OQ-45.2 Validation Subscale Covariances (*continued*)

			Estimate	S.E.	C.R.	P
Depression	↔	Anxiety	0.395	0.04	8.917	.0001
Danger	↔	Conflict	0.186	0.03	6.406	.0001
Alcohol	↔	Conflict	0.002	0.01	0.474	.636
LoveRelat	↔	Conflict	0.327	0.05	6.813	.0001
Somatic	↔	Conflict	0.288	0.04	7.002	.0001
Role	↔	Conflict	0.258	0.04	6.097	.0001
Anxiety	↔	Conflict	0.307	0.04	7.332	.0001
Depression	↔	Conflict	0.313	0.04	7.722	.0001
Alcohol	↔	Relationships	-0.006	0.01	-1.18	.238
LoveRelat	↔	Relationships	0.217	0.04	5.506	.0001
Somatic	↔	Relationships	0.151	0.03	4.848	.0001
Role	↔	Relationships	0.291	0.04	6.899	.0001
Anxiety	↔	Relationships	0.158	0.03	4.708	.0001
Depression	↔	Relationships	0.256	0.04	7.295	.0001
LoveRelat	↔	Danger	0.17	0.03	5.63	.0001
Somatic	↔	Danger	0.236	0.03	7.113	.0001
Role	↔	Danger	0.3	0.04	7.842	.0001
Anxiety	↔	Danger	0.297	0.04	7.984	.0001
Depression	↔	Danger	0.277	0.03	8.065	.0001
Somatic	↔	Alcohol	-0.004	0.01	-0.91	.362
Role	↔	Alcohol	-0.01	0.01	-1.53	.127

Table F3: OQ-45.2 Validation Subscale Covariances (continued)

			Estimate	S.E.	C.R.	P
Anxiety	↔	Alcohol	-0.01	0.01	-1.92	.055
Depression	↔	Alcohol	-0.005	0.01	-1.07	.283
Role	↔	LoveRelat	0.258	0.05	5.516	.0001
Anxiety	↔	LoveRelat	0.284	0.04	6.429	.0001
Depression	↔	LoveRelat	0.429	0.05	8.116	.0001
Anxiety	↔	Somatic	0.392	0.05	8.104	.0001
Depression	↔	Somatic	0.347	0.04	8.039	.0001
Depression	↔	Role	0.427	0.05	9.009	.0001

Table F4: OQ-45.2 Validation Correlations

			Estimate
Relationships	↔	Conflict	.485
Danger	↔	Relationships	.603
Alcohol	↔	Danger	-.054
LoveRelat	↔	Alcohol	-.081
Somatic	↔	LoveRelat	.445
Role	↔	Somatic	.777
Anxiety	↔	Role	.674
Depression	↔	Anxiety	.742
Danger	↔	Conflict	.633

Table F4: OQ-45.2 Validation Correlations (*continued*)

			Estimate
Alcohol	↔	Conflict	.03
LoveRelat	↔	Conflict	.696
Somatic	↔	Conflict	.717
ROle	↔	Conflict	.478
Anxiety	↔	Conflict	.669
Depression	↔	Conflict	.696
Alcohol	↔	Relationships	-.08
LoveRelat	↔	Relationships	.482
Somatic	↔	Relationships	.393
Role	↔	Relationships	.565
Anxiety	↔	Relationships	.359
Depression	↔	Relationships	.596
LoveRelat	↔	Danger	.477
Somatic	↔	Danger	.773
Role	↔	Danger	.735
Anxiety	↔	Danger	.853
Depression	↔	Danger	.813
Somatic	↔	Alcohol	-.051
Role	↔	Alcohol	-.085
Anxiety	↔	Alcohol	-.106
Depression	↔	Alcohol	-.056

Table F4: OQ-45.2 Validation Correlations (continued)

			Estimate
Role	↔	LoveRelat	.393
Anxiety	↔	LoveRelat	.509
Depression	↔	LoveRelat	.786
Anxiety	↔	Somatic	.823
Depression	↔	Somatic	.744
Depression	↔	Role	.683

Table F5: OQ-45.2 Validation Variances

	Estimate	S.E.	C.R.	P
Depression	0.521	0.06	8.267	.0001
Anxiety	0.543	0.07	7.644	.0001
Role	0.75	0.08	9.127	.0001
Somatic	0.417	0.07	5.928	.0001
LoveRelat	0.572	0.1	5.862	.0001
Alcohol	0.017	0	9.114	.0001
Danger	0.223	0.04	5.695	.0001
Relationships	0.354	0.07	5.217	.0001
Conflict	0.387	0.07	5.975	.0001
e02	0.729	0.05	13.93	.0001
e03	0.595	0.04	14.08	.0001

Table F5: OQ-45.2 Validation Variances (*continued*)

	Estimate	S.E.	C.R.	P
e05	0.694	0.05	14.02	.0001
e06	0.431	0.04	9.951	.0001
e08	0.552	0.04	13.93	.0001
e09	0.5	0.04	11.67	.0001
e10	0.606	0.05	12.63	.0001
e11	0.02	0	13.91	.0001
e15	0.527	0.04	13.33	.0001
e22	0.574	0.04	13.16	.0001
e23	0.586	0.04	13.44	.0001
e24	0.548	0.04	13.89	.0001
e25	0.859	0.06	13.48	.0001
e27	0.905	0.07	13.89	.0001
e29	0.768	0.06	13.75	.0001
e31	0.495	0.04	13.51	.0001
e33	0.643	0.05	13.23	.0001
e34	0.991	0.07	14.27	.0001
e35	0.423	0.03	14.58	.0001
e36	0.661	0.05	13.27	.0001
e40	0.562	0.07	8.483	.0001
e42	0.508	0.04	13.34	.0001
e45	0.968	0.07	14.04	.0001

Table F5: OQ-45.2 Validation Variances (*continued*)

	Estimate	S.E.	C.R.	P
e01	0.418	0.03	12.73	.0001
e07	1.256	0.09	13.47	.0001
e18	0.69	0.06	10.79	.0001
e19	0.709	0.06	12.29	.0001
e20	0.642	0.05	14.25	.0001
e26	0.005	0	2.651	.008
e30	0.403	0.06	6.807	.0001
e37	0.806	0.07	12.45	.0001
e43	0.601	0.04	14.25	.0001
e04	0.691	0.05	13.39	.0001
e21	0.854	0.06	14.45	.0001
e28	0.534	0.05	10.18	.0001
e32	0.009	0	7.789	.0001
e39	0.643	0.05	14.32	.0001
e44	0.771	0.06	14.14	.0001
e17	1.383	0.1	13.43	.0001
e38	0.486	0.05	10.72	.0001
e13	0.524	0.04	14.26	.0001
e16	1.143	0.08	14.09	.0001

VITA

Jason E. Chapman was born in Portsmouth, Ohio, on April 20, 1976, and moved to Easley, South Carolina at the age of 12. He completed his undergraduate work at Clemson University in SC, and he graduated *Magna Cum Laude* with a Bachelor of Arts degree in psychology and sociology in 1998. That fall he began a doctoral program in Clinical & Health Psychology at MCP-Hahnemann University, where he received his Master of Arts degree in June of 2000. While at MCP-Hahnemann (Drexel), he worked as assessor, coordinator, and ultimately co-principle investigator in a program evaluation project of a supportive housing program for previously homeless, chronically mentally ill individuals. During this time, he began his clinical work and completed a two year practicum in private practice at the Center for Cognitive and Behavior Therapy in Wilmington, DE. Also while at MCP-Hahnemann (Drexel), he co-taught graduate statistics at LaSalle University and worked as a statistics, database design, and research consultant in multiple settings. In July of 2002, he began his pre-doctoral internship at Geisinger Medical Center in Danville, PA, where he completed rotations in outpatient and primary care psychotherapy, emergency services and hospital consultation, inpatient psychiatry, and physical rehabilitation/ neuropsychology and assisted in the implementation of a patient-focused research project. Following graduation in 2003, he will begin a postdoctoral fellowship under the direction of Aaron T. Beck in the Psychopathology Research Unit at the University of Pennsylvania where he will be involved in a variety of research and clinical activities related to cognitive therapy of borderline personality disorder, repeat suicide attempters, and schizophrenia.